4-11-2019

Investigation of Cold Stress Induced Disease Resistance (SIDR) in Grapevines and Arabidopsis

Cal D. Palumbo
cp1015@rit.edu

Investigation of Cold Stress Induced Disease Resistance (SIDR) in Grapevines and *Arabidopsis*

By Cal D. Palumbo

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of
Master of Science in Bioinformatics

Thomas H. Gosnell School of Life Sciences
College of Science
Rochester Institute of Technology
Rochester NY
April 11th, 2019

Rochester Institute of Technology
Thomas H. Gosnell School of Life Sciences
Bioinformatics Program

To: Head, Thomas H. Gosnell School of Life Sciences

The undersigned state that Cal D. Palumbo, a candidate for the Master of Science degree in Bioinformatics, has submitted his thesis and has satisfactorily defended it.

This completes the requirements for the Master of Science degree in Bioinformatics at Rochester Institute of Technology.

Thesis committee members:

Name	**Date**
Michael V. Osier, Ph.D. Thesis Advisor	
Lance Cadle-Davidson, Ph.D.	
Dawn Carter, Ph.D.	

Feng Cui, Ph.D.
Director of Bioinformatics MS Program

475-4115 (voice)
fxcsbi@rit.edu

Acknowledgments

First, I would like to thank my advisor, Dr. Michael Osier, for his advice and teachings. Your guidance has been indispensable, not only while working on this project, but throughout my entire time in this program.

I would like to thank my collaborator, Bill Weldon, who performed the sample collection and sample phenotyping for this project.

A special thank you to my thesis committee member, Dr. Lance Cadle-Davidson, for all of his teachings and encouragement.

I would also like to thank my thesis committee member, Dr. Dawn Carter, for supporting me during this project and providing resources to fill in the gaps in my knowledge.

Table of Contents

List of Figures

List of Tables

Abstract

The focus of this thesis project was to investigate the impact of cold temperature conditions on the resistance of *Vitis vinifera* (grapevines) to powdery mildew, a phenomenon known as cold stress-induced disease resistance (SIDR). The model organism *Arabidopsis thaliana* was used to determine specific defense mechanisms of plant-pathogen resistance. An RNA-Seq time course experiment was performed for the two organisms: *V. vinifera* and *A. thaliana*. The time-series datasets consisted of data points where samples were exposed to an acute (less than 4 hours) cold (below 8°C) temperature for set times prior to inoculation with powdery mildew. The acute cold time points used ranged from 48 hours, 36 hours, 24 hours, and 12 hours prior to inoculation. An untreated control group, which was not exposed to any acute cold treatment, was used to compare between the treatments. The significant, differentially expressed genes were evaluated and mapped to the pathways of the respective organisms.

The outcome of this project was the identification of pathways, as well as potential genes of interest, involved with cold stress-induced disease resistance. There has been limited research on genetic mechanisms of cold stress-induced disease resistance. This project provides an improved understanding of the interactions between host stress and the epidemiology of the biotrophic pathogen powdery mildew.

Introduction

According to the U.S. Department of Commerce (2017) liquor store sales of beer and wine contribute more than $50 billion to the U.S. economy. This dollar amount does not consider the indirect economic impacts of the wine industries. The American wine industry is responsible for the combined direct and indirect impact estimated at more than $200 billion (WineAmerica, 2017). The primary agricultural product used in wine production is grapes. In 2013, the U.S. produced more than 7 million tons of grapes (Wineamerica, 2014). While wine can be produced using any grape varieties in the *Vitis* genus, the most desired grapes come from the grapevine species *Vitis vinifera*.

Grapevines need to cope with a number of environmental stresses as well diseases and pests in order to thrive. Of the possible pathogens, the fungal disease powdery mildew (PM) is considered among the most important diseases in the world for grapes (Gent et al., 2009), (Moyer, 2011). The grapevine species of powdery mildew, *Erysiphe necator*, is native to eastern North America. In 1845 the disease was introduced to European vineyards where it caused extensive losses as it spread rapidly throughout the continent (Wilcox, 2003). Since these times control of powdery mildew has been a serious concern. Left unchecked, powdery mildew can wipe out an entire crop on its own or make the crop susceptible to other pathogens or abiotic stresses.

Powdery Mildew Overview

Powdery mildew is an obligate biotrophic fungus meaning that it can only grow on living plant tissue. A number agricultural crops are affected by powdery mildews, including artichoke, beans, beets, carrot, cucumber, eggplant, lettuce, melons, parsnips, peas, peppers, pumpkins, radicchio, radishes, squash, tomatillo, tomatoes, and turnips (UC-IPM, 2008).

Powdery mildews are caused by a broad range of genera and species residing in the order Erysiphales in the Fungi kingdom. Most powdery mildew species are host-specific, meaning that each susceptible plant can only be infected by one (or few) species of powdery mildew. Grapevines in the species *V. vinifera* are infected by the powdery mildew species *Erysiphe necator* (syn *Uncinula necator*). Symptoms of powdery mildew include white fungal growth on berries and leaves and dark lesions on infected vine (Figure 1).

Figure 1: Powdery mildew (Erysiphe necator) shown on grapes.

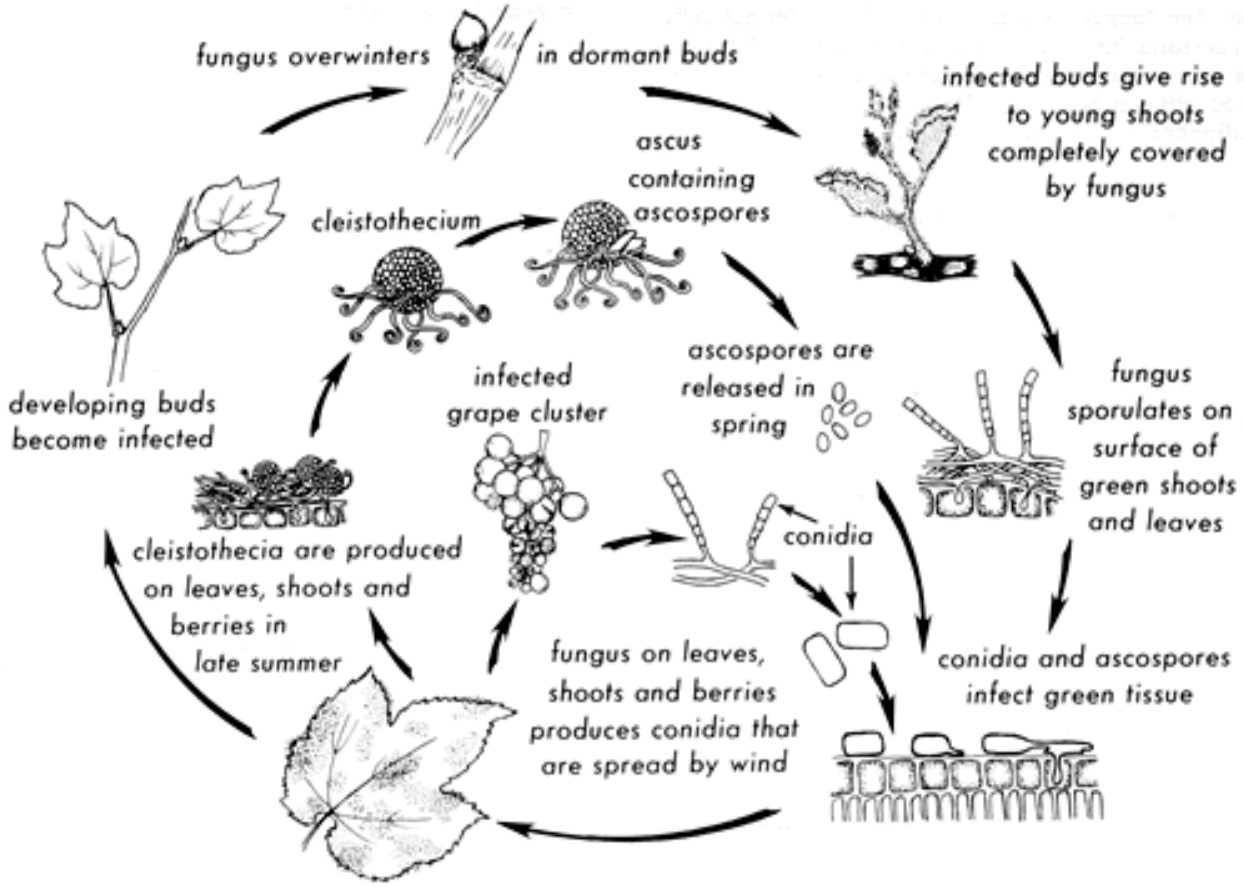
(Photo: L. J. Bettiga)

Life Cycle of Powdery Mildew

The disease lifespan of powdery mildew follows a yearly cycle (Figure 2). During the late summer, powdery mildew colonies begins to produce chasmothecia. Chasmothecia contain ascospores which are the result of mating among individual powdery mildew colonies

(Martinson & Wilcox, 2013). These hardened structures are responsible for initiating the disease process during the start of the growing season in spring. Starting in the Spring, as the temperature rises above 14-16° C, chasmothecia release ascospores which begin to infect nearby green tissue. As the fungal colony grows, it produces conidia (spores) as a form of asexual reproduction. These conidia are dispersed via airflow to surrounding plants. When the conidia make contact with a suitable host the infection process begins. According to Micali et al. (2008), once a conidium/spore has landed on a viable host, this infection process begins in as little as 1 hour. As the fungi colony grows, it continues to produce conidia which continue to spread the infection through the summer months.

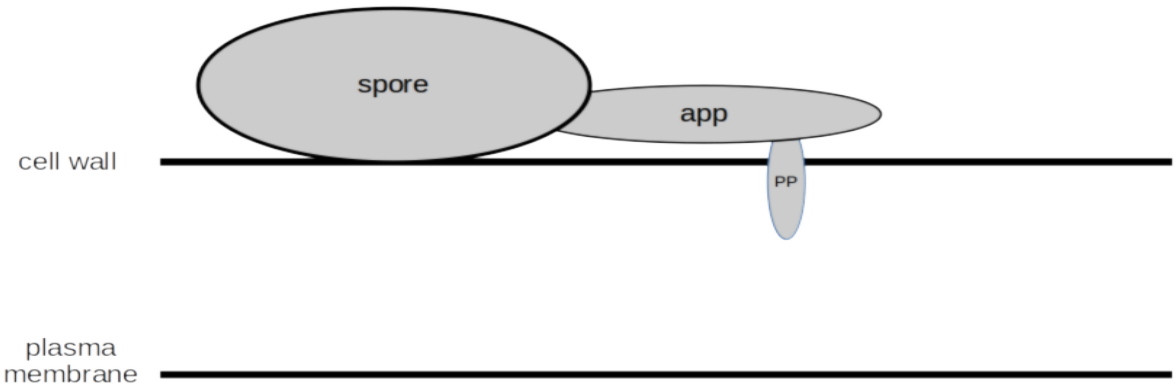
Figure 2: Grapevine powdery mildew (Erysiphe necator) life cycle.



(Drawing by R. Sticht.)

Powdery Mildew Infection and Host Response

A set of complex cellular processes are deployed by both the plant host and the obligate biotroph powdery mildew during infection. When a powdery mildew conidium has landed on a viable host, and the infection process has begun, the fungus will create an appressorium, the primary infection structure. The newly formed appressorium will penetrate into the host cell wall, forming a haustorium, a specialized, intracellular structure, within the epidermal layer of the host plant (Figure 3). The haustorium will serve as the active interface between the fungus and the host plant. The fungus will retrieve necessary nutrients from the host cells, while secreting proteins to counteract the host defense response.



app:appressorium, pp:penetration peg

Figure 3: Powdery mildew spore with appressorium formation

Plants have two major methods for disease resistance, basal defense and R-gene mediated defense. The basal defense represents the first line of defense for the host. Basal defenses look to be triggered by cell wall components which are released by the hydrolytic activity of enzymes secreted by the appressorium, as well as common features of the fungus,

known as pathogen-associated molecular patterns (PAMPs) (Gururani et al., 2012). To overcome this basal defense, the powdery mildew haustorium produces "effector" proteins which are used to interrupt the activation of host plant defenses. Continuing this veritable arms race, plants have developed defense responses known as effector triggered immunity (ETI). ETI responses in plants are controlled by resistance (R) genes which encode proteins that interact with the effector proteins, leading to defense responses in the host (Qiu, Feechan, & Dry, 2015). Further, knockdown of susceptibility genes such as Mildew Locus-o (*MLO*) through RNA interference can reduced powdery mildew severity by up to 77% (Pessina et al. 2016).

Control Methods

In commercial vineyards, the control of powdery mildew generally requires the significant use of fungicides, as well rigorous canopy management (Gadoury et al., 2012). Fungicides are used to reduce the spread of infection. Fungicide applications are determined by advisory systems that use a risk index for the vineyards (Bendek et al., 2007). To calculate the risk of disease development for a vineyard, air temperature is used as input for the prediction model. In the state of California, over 15 million pounds of chemicals pesticides were used on grape crops in 2012 (Kegley et al. 2016). The continued use of pesticides could have dangerous unseen side effects. Most commonly used fungicides in vineyards consist of various sulfur preparations. Exposure to sulfur can lead to irritation of the skin, eyes, and respiratory tract. This can be hazardous to people in the close proximity as well as to the environment (Youakim, 2006), (Raanan et al., 2017). On crops sensitive to sulfur or when several diseases need simulataneous control, site-specific fungisides are commonly used, which after extensive use can also lead to resistance in the pathogen. The susceptible pests are controlled, but host resistant individuals of the same species reproduce and increase in absence of competition (Gent et al.,

2009). Over time, the resistant strains become the prevalent population spreading the infection. In the control of *E. necator*, conventional management with modern, organic fungicides has been compromised on several occasions since 1980 by the evolution of fungicide resistance (Gadoury et al., 2012). This makes host-resistance to PM a valuable and desirable trait. Grape varieties with resistance to powdery mildew are currently being developed, using either conventional or transgenic approach (Fuller, Alston, & Sambucci, 2014).

Stress Induced Disease Resistance

Plants are subject to two main types of deleterious stresses during their lives, biotic and abiotic (Petrov, Hille, Mueller-Roeber, & Gechev, 2015). An example of biotic stress is the plant's intersection with the biotroph powdery mildew. Abiotic stresses include extreme temperatures, high salinity, excessive light, water deprivation, pollutants such as ozone and herbicides, high concentrations of heavy metals, and excessive UV radiation (Petrov et al. 2015). There is a phenomenon during powdery mildew infection where environmental stresses can have a negative impact on the development of the fungus. The fungus develops on the outside of the plant tissue making it susceptible to extreme conditions of the external environment. Recently, this phenomenon has been observed in *V. vinifera* during acute cold events. These acute cold events ranged from 5 minutes to 8 hours in duration with temperatures below 8 ºC (Moyer et al., 2010). Following these acute exposures, there was observed: "death of hyphal segments, and a prolonged latency" associated with the infection (Moyer et al., 2016). This observed disease resistance, however, was temporary. The resistance response diminished to basal levels within 48 hours following exposure. Moyer et al. (2010) proposed that the exposure of grape leaf tissue to extreme temperatures either made host tissue unsuitable for colonization or activated a temporary host defense response.

Phenotyping results noted in Weldon et al., (manuscript in preparation) showed a significant difference in the percent susceptibility of grapevine sample when exposed to an acute cold treatment at 24 hours prior to inoculation with powdery mildew. In the same study, there was a significant difference shown in percent susceptibility in the species *Arabidopsis thaliana.* The percent susceptibility was reduced for samples exposed to an acute cold treatment at both 12 and 24 hours prior to inoculation with powdery mildew.

Moyer et al. (2016) presented three potential hypotheses for the cause of the cold-SIDR response. The first hypothesis involved the cold temperature impacting photosynthesis efficiency. The second hypothesis is that basic physiological responses to cold that mimic a short term induction of "ontogenic resistance" type response. The short term response would include a decrease in carbon assimilation and vegetative growth rate as well as an increase in calcium signaling and reactive oxygen species (ROS) generation. The third hypothesis involved the regulation of plant hormones signaling, Gibberellin biosynthesis and DELLA proteins and the abscisic acid (ABA) pathway. Determining the exact host response is difficult though because of the obligate biotrophic nature of *E. necator*. In vitro studies of the biochemical pathways that lead to stress responses are challenging because those pathways can also be affected by the powdery mildew infection. To assist in this task the use of a model organism is necessary.

Arabidopsis thaliana Overview

Arabidopsis thaliana is a small flowering plant that is a member of the mustard (*Brassicaceae*) family. *A. thaliana* is used widely in biological sciences and there are extensive genetic maps of the organism's chromosomes. In addition, the rapid life cycle of the plant makes it a model organism. The biological pathways of the organism have also been broadly annotated. Though *A. thaliana* follows the trend of only being susceptible to a certain species of powdery

mildew, mutations in certain “non-host” genes allow it to be infected by other powdery mildew species. Specifically, mutations in the PEN1 and PEN1/3 non-host genes make *A. thaliana* susceptible to infection by *E. necator*. This susceptibility makes *A. thaliana* the model organism to use when investigating host defenses against grapevine powdery mildew.

RNA-Seq Analysis

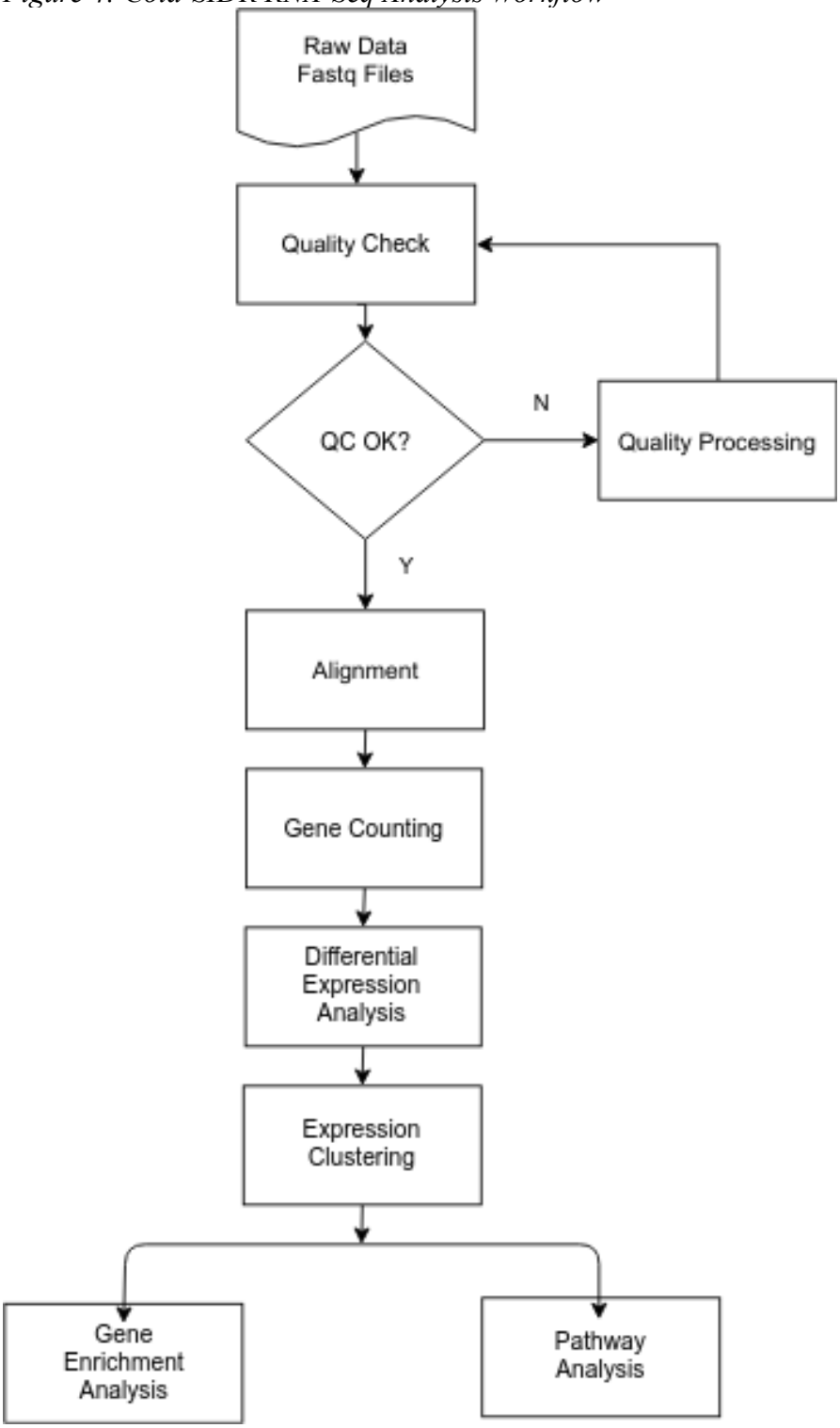
High Throughput RNA sequencing (RNA-Seq) is a valuable technique for monitoring gene expression (Wang, Gerstein, & Snyder, 2009). RNA-Seq utilizes mRNA transcripts to provide deep insight into an organism's gene expression. With tradition RNA-Seq method, the isolated mRNA transcript is randomly fragmented and converted into a cDNA library. The cDNA fragments are then sequenced using a next-generation sequencing technology. Gene expression is proportional to the total number of reads (cDNA fragments sequenced) corresponding to a given transcript of a gene. A limitation of traditional RNA-Seq is that longer transcripts will produces more fragments/reads than shorter transcripts (Tandonnet & Torres, 2016). The 3' RNA-Seq method, overcomes this limitation by retaining only one fragment per transcript from the 3′ region. Using this strategy, the expression levels are estimated directly by the number of reads corresponding to a single gene.

The focus of this thesis project was to test the hypothesis that the genes and gene pathways mentioned in the cold-SIDR review (Moyer et al., 2016) will have differential expression patterns in the treatment groups that correlate with the observed transient resistance phenotype. An RNA-Seq analysis was performed to investigate the differentially expressed genes in each species, with the end goal of comparing and contrasting those genes between the organisms.

Methods and Materials

Analysis Pipeline Overview

Figure 4: Cold-SIDR RNA-Seq Analysis Workflow



Dataset Creation

To determine the gene homologs and pathways that contribute to the cold stress-induced disease resistance (SIDR) in the two organisms (*V. vinifera*, *A. thaliana*) two datasets were obtained. The datasets were created by William Weldon using the methods described in Weldon et al., manuscript in preparation. An overview of the datasets are described below.

V. vinifera Dataset Creation

An experiment started with 15 leaves, with each leaf being subsampled into 5 disks (1 cm in diameter), assigned to each of 5 treatment time points in total: untreated control (UTC), 12 hours post cold (hpc), 24 hpc, 36 hpc, and 48 hpc. Each treatment group contained fifteen leaf disks, one disk from each leaf, positioned in one of fifteen positions on a petri dish (Figure 5). The position of the leaf disk was maintain between all treatment petri dishes because leaf disks from the same leaf are known to have similar resistance responses, which is a component of variability that can be accounted for statistically. The time points reflected the hours since being cold treated for 4hrs at 4°C/39.2°F. From these treated leaf disks five disks were removed and sent for RNA sequencing. The remainder of the leaf disks were inoculated with powdery mildew (*Erysiphe necator*). Figure 6 shows the timeline for the experiment. This experiment was repeated a total of 3 times. The phenotypic responses for each treatment group were recorded to determine the percent susceptibility of each group to powdery mildew relative to the untreated control (UTC) group.

Figure 5: Vitis vinifera Disk Placement

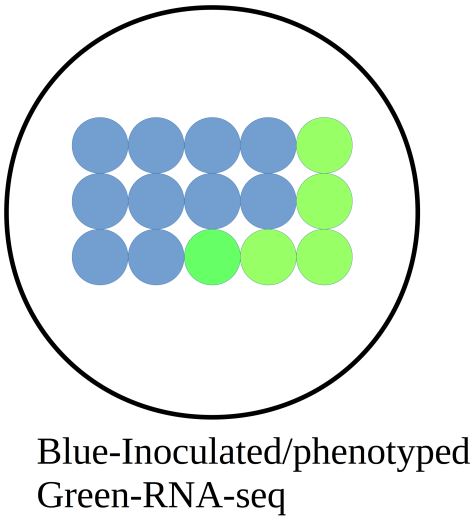
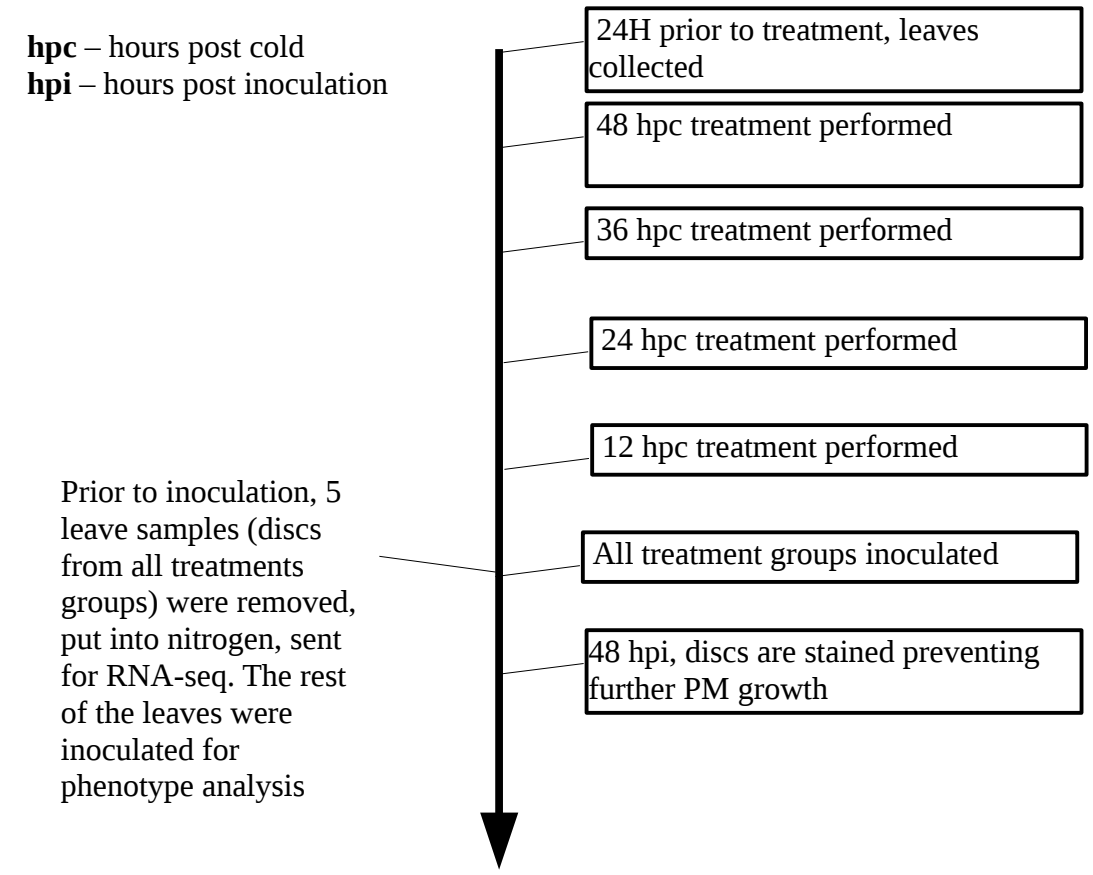


Figure 6: Vitis vinifera Sample Treatment Timeline



A. thaliana Dataset Creation

Two lines of *A. thaliana* were used; wild-type Col-0 and a PEN1 mutant in Col-0 background. Each line followed the same experimental design. An experiment started with 8 seedlings per line per treatment. There were 5 treatment time points in total: untreated control (UTC), 12 hours post cold (hpc), 24 hpc, 36 hpc, and 48 hpc, reflecting the hours since the initiation of cold treatment for 4hrs at 4°C (39.2°F) To reduce variability due to environment, both lines were grown in the same petri dish for a given treatment in a split plot design (Figure 7). From the seedlings, 3-4 were removed and sent for RNA sequencing. The remaining seedlings were inoculated with powdery mildew (*E. necator*). The phenotypic responses for each treatment group were recorded to determine the percent susceptibility of each group to powdery mildew relative to the untreated control (UTC) group. Figure 8 shows the treatment timeline for the experiment. This experiment was repeated a total of 3 times.

Figure 7: Arabidopsis thaliana Seedling Placement

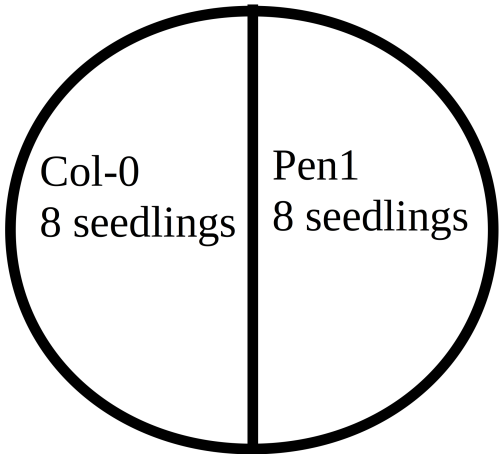
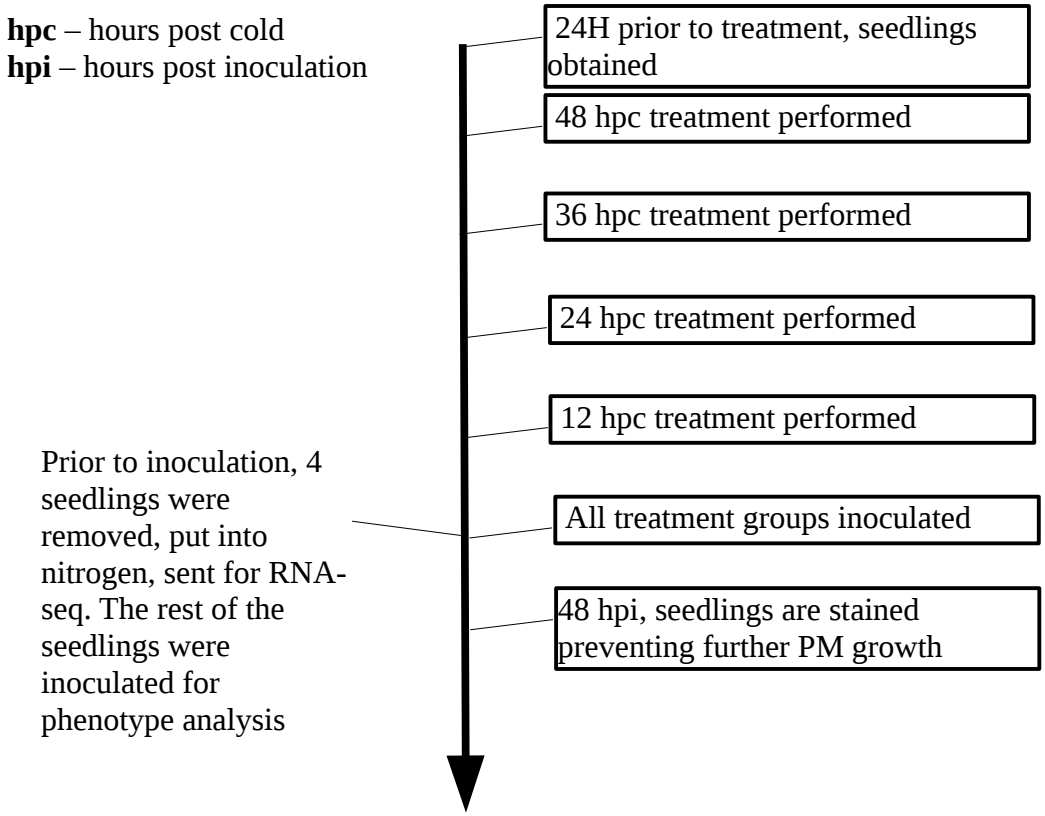


Figure 8: Arabidopsis thaliana Sample Treatment Timeline



RNA-Seq Library Construction

The datasets for *V. vinifera* and *A. thaliana* were generated using 3'RNA-Seq library construction. Both libraries were generated using the QuantSeq 3' mRNA-Seq Library Prep Kit for Illumina. Sequencing then took place using an Illumina NextSeq500 with an anticipated read length of 75 bp. The library construction and sequencing steps were performed by the Cornell University Biotechnology Resource Center, Ithaca, NY.

Preprocessing

A crucial step in the RNA-Seq pipeline is the initial quality control and preprocessing of the RNA sequence datasets. Quality problems including low-confidence bases, sequence-

specific bias, polymerase chain reaction artifact, untrimmed adaptors, and sequence contamination may need to be addressed (Korpelainen et al., 2014). First, the program FastQC (Andrews, 2010) was used to access the quality controls (QC) with regards to the sequence read files. FastQC returns results pertaining to various aspects of sequencing data quality in a series of eleven modules (Figure 9).

Figure 9: FastQC Output Modules

Basic statistics	A summary description of the reads in the file, including file name, assumed quality score encoding scheme, number of sequences, sequence lengths, and GC content.
Per base sequence quality	A plot representing the distributions of the quality scores for each nucleotide position for the set of reads, depicted as box-and-whisker plots for each position.
Per sequence quality scores	A plot describing the distribution of mean sequence quality scores. The x-axis holds mean sequence quality scores while the y-axis represents the number of reads having that mean quality score.
Per base sequence content	A plot of the percent content of each nucleotide for each position within the read length over the whole set of reads.
Per sequence GC content	A plot of the mean GC content distribution over the set of reads. The x-axis represents the range of possible mean GC content percent values, while the y-axis provides the number of reads with each mean GC content.
Per base N content	A plot of the percent "N" (indeterminate nucleotide) over the entire set of reads by nucleotide position.
Sequence Length Distribution	A plot of the lengths of the reads in the set. The x-axis shows the range of read lengths seen, while the y-axis provides the number of reads with each length.
Sequence Duplication Levels	A plot of the percent of reads duplicated at a given level (y-axis) against the range of levels of duplication (x-axis).
Overrepresented sequences	A list of any reads found to each compose greater than 0.1% of the entire set of reads, if any do so.
Adapter Content	A plot of the percent of reads in which established adapter sequences have been detected (y-axis) against nucleotide position within the read (x-axis).
Kmer Content	A plot and list of those 7-mers found to have a significant location bias in reads.

(FastQC documentation)

The FastQC results for the 90 samples of each dataset were parsed using a custom PERL script.

The initial run of FastQC on the *V. vinifera* dataset showed that of the 90 sequence files, 3 prompted warnings for a check of overrepresented sequences. As described in Table 1, this meant that those files contained sequences which were found to represent more than 0.1% of

the total number of sequences in each file, which may be reasonable for expression data. There were also 78 files which prompted warnings due to the GC content of their sequences. For the *A. thaliana* dataset, the initial run of FastQC produced several failures, as well as several warnings. Of the 90 sequence files, 78 proposed failures for the GC content of their respective sequences. There were also 90 sequence files which prompted warning for overrepresented sequences. The overrepresented sequence that caused the failure was composed of poly-G (Guanine) nucleotides. An explanation for the poly-G sequences can be deduced by evaluating the sequencing method used. The NextSeq platform uses a two-color chemistry system, where a "G" nucleotide is designated by the color "black" (no color). When the nucleotide color signal is too weak to detect, it is determined to have no color and will be recognized as a "G" in the base calling stage (Chen, 2018).

To rule out the possibility of contamination or sample bias, these warnings were investigated further. According to the NCBI, the *V. vinifera* genome assembly had a median GC content of 33.75%, while the *A. thaliana* genome assembly had median GC content of 36.15%. The 90 samples from the *V. vinifera* dataset had a median GC content of 39%. The 90 samples from the *A. thaliana* dataset had a median GC content of 37%. Since the datasets consisted of transcriptome sequencing, it was possible that the GC content of the transcriptome varied from the GC content of the genome. Singh, Ming, and Yu (2016) found that the GC content for the coding sequences (CDS) in *V. vinifera* was ~44.5% while the GC content for the CDS in *A. thaliana* was ~44.1%. Since this article's publication, new genome annotations for *V. vinifera* (Canaguier et al., 2017) and *A. thaliana* (Cheng et al., 2017) were released. To account for the possible changes in the new genome annotations, the GC content for each annotation was calculated using BEDTools (Quinlan and Hall, 2010). The GC content for the CDS in new

genome annotation for *V. vinifera* was ~38.8%. The GC content for the CDS in new genome annotation for *A. thaliana* was ~42.1%.

To address the quality issues of the *V. vinifera* dataset a number of tools from the Fastx-Toolkit were used. At the recommendation of the sequencing facility, the first 13 bases were removed from the 5' end of the sequence read using the Fastx Trimmer tool as this region contained sequencing errors due to random priming. The FastX Quality Trimmer tool was used to remove low quality bases from the 3' end of each read. The minimum Phred quality score used to initiate base trimming/removal was 20. Sequence reads shorter than 38 bases were remove from the analysis. The FastX Quality Filter tool was used to remove sequences reads with a low overall quality. This quality filtering was processed in two stages. The first stage removed sequences reads which failed the requirement of having 80% of the bases in the read above a Phred quality score of 20. The second stage removed sequences reads that failed the requirement of having 100% of the bases in the read above a Phred quality score of 13. The settings for these tools were selected to optimize the quality of the sequence files while maintaining a high number of sequence reads within each file. The same set of tools and settings were also used to address the quality issues of the *A. thaliana* dataset, with the addition of the FastX Artifact Filter to remove sequencing artifact. A description for the tools and their respective settings for each organism are listed in Table 1.

V. vinifera

Tool	Setting Flag	Description
FastX Trimmer	-f 13	Trim the first 13 bases from 5' end of read
FastX Quality Trimmer	-t 20	Quality threshold – nucleotides with lower quality will be trimmed (from the end of the sequence)
FastX Quality Trimmer	-l 38	Minimum length - sequences shorter than this (after trimming) will be discarded.
Fastq Quality Filter Part 1	-q 20	Minimum quality score to keep. (Phred Score = 20)
Fastq Quality Filter Part 1	-p 80	Minimum percent of bases that must have [-q] quality (≥ 80% of bases with Phred score > 20)
Fastq Quality Filter Part 2	-q 13	Minimum quality score to keep. (Phred Score = 13)
Fastq Quality Filter Part 2	-p 100	Minimum percent of bases that must have [-q] quality (100% of bases with Phred score > 13)

A. thaliana

Tool	Setting Flag	Description
FastX Artifact Filter	Default Settings	Filters sequencing artifact (reads with all but 3 identical bases)
FastX Trimmer	-f 13	Trim the first 13 bases from 5' end of read
FastX Quality Trimmer	-t 20	Quality threshold – nucleotides with lower quality will be trimmed (from the end of the sequence)
FastX Quality Trimmer	-l 38	Minimum length - sequences shorter than this (after trimming) will be discarded.
Fastq Quality Filter Part 1	-q 20	Minimum quality score to keep. (Phred Score = 20)
Fastq Quality Filter Part 1	-p 80	Minimum percent of bases that must have [-q] quality (≥ 80% of bases with Phred score > 20)
Fastq Quality Filter Part 2	-q 13	Minimum quality score to keep. (Phred Score = 13)
Fastq Quality Filter Part 2	-p 100	Minimum percent of bases that must have [-q] quality (100% of bases with Phred score > 13)

Table 1: FastX Quality Control Tools and Settings

Alignment

For both datasets, an alignment was performed between the sequence files and a reference genome. For *V. vinifera*, reads were aligned to the 12xV2 genome and the Vcost.v3 annotation (Canaguier et al., 2017). For *A. thaliana*, reads were aligned to the Tair10 genome (Berardini et al., 2015) using the Araport11 annotation (Cheng et al., 2017). The program STAR (Spliced Transcripts Alignment to a Reference) (Dobin et al., 2013) was used for this alignment step. The settings used for the STAR are described in Table 2.

Flag	Setting	Description
--runThreadN	6	Specify the number of threads
--outFilterMultimapNmax	10	Specify the max number of multiple alignments allowed for a read: if exceeded, the read is considered unmapped
--outFilterMismatchNoverLmax	0.04	Specify the max number of mismatches per pair relative to read length: for 2x100b, max number of mismatches is 0.04*200=8 for the paired read
--outFilterIntronMotifs	RemoveNoncanonicalUnannotated	Filter out alignments that contain non-canonical unannotated junctions when using annotated splice junctions database. The annotated non-canonical junctions will be kept.
--readFilesCommand	gunzip -c	Uncompresses the input files and sends output to STAR as input

Table 2: STAR Alignment Settings

The STAR software output the aligned reads in a sequence alignment map (.sam) file format. This output was converted to a binary alignment map (.bam) file format using the command-line

program SAMtools (Li et al., 2009). The settings used for the SAMtools are described in Table 3.

V. vinifera

Setting Flag	Description
sort -n	Sort alignments by read name
-@ 2	Number of additional threads to use

A. thaliana

Setting Flag	Description
sort	Sort alignments by leftmost coordinates
-@ 6	Number of additional threads to use

Table 3: SAMtools settings used for post alignment processing

Gene Counting

In an RNA-Seq experiment, it is necessary to know how many reads fall within the exonic regions of each gene. The python script HTSeq (htseq-count) v0.9.1 was used to determine these gene counts (Anders, Pyl, & Huber, 2015). The datasets were separated by their respective treatments and were processed in parallel on the server. The following settings were used for both the *V. vinifera* and *A. thaliana* datasets (Table 4).

Flag	Setting	Description
--stranded	no	Whether the data is from a strand-specific assay
--nounique	none	If the read is associated with more than one feature, the read is counted as ambiguous and not counted for any features.
--order	name	Alignment files have been sorted by read name

Table 4: HTseq (htseq-count) Settings

Differential Expression

A differential expression analysis was performed separately for each of the datasets. The R package DESeq2 (Love, Huber, & Anders, 2014) was used with the output from HTSeq. Metadata about the samples were extracted from their filesnames using regular expression. Using the function *DESeqDataSetFromHTSeqCount*, the gene counts and metadata for the

samples were combined into a *DESeqDataSet* object. The *DESeqDataSet* also contained an associated design formula, which is used to model the gene counts.

Three design formulas (models) were used for the *V. vinifera* dataset. Each design formula was meant to identify specific genes of interest which followed a certain expression pattern. All three designs included a "Disk Number" term in the formula. This term was meant to control for leaf-to-leaf variability, as one disk from each leaf was assigned to each treatment and tracked. Because a leaf could not be used in multiple independent experiments, experiment number was confounded with "Disk Number", so was not used in the models. The "Circadian Rhythm" term in the formula #2 was meant to account for variability due to the time of day when the treatments were performed. This was a binary variability where the UTC, 24hpc, 48hpc treatments were classified as day and the 12hpc, 36hpc were classified as night. The "Percent Susceptibility" term was meant to identify genes with counts that could be explained by changes in the percent susceptibility of the treatment group. This percent susceptibility was determined from results from the sample phenotyping. The "Treatment Time" and "Treatment Time^2" terms were meant to identify genes that fit the pattern of having a peak up or down regulation at the 24hpc treatment group. The "Treatment" term was meant to identify genes that show change in expression across the different treatment groups. This term was a categorical (factor) variable with the UTC treatment serving as the reference group. In DESeq2, a likelihood ratio test (LRT) was performed to compare how well the count data for a gene fit a "full model" (design formula with all variables) compared to a "reduced model," with the variable(s) of primary interest removed. The full and reduced models use for the *V. vinifera* analysis are illustrated in Table 5.

#	Type	Design Formula	Explanation
1	Full	~ Disc Number + Treatment	Identify genes which changed between the treatment groups, while controlling for leaf variability
	Reduced	~ Disc Number	
2	Full	~ Disc Number + Circadian Rhythm + Percent Susceptibilit	Identify genes which are explained by a change in % susceptibility, based on phenotypic observations, while controlling for leaf variability and circadian rhythm
	Reduced	~ Disc Number + Circadian Rhythm	
3	Full	~ Disc Number + Treatment Time + Treatment Time^2	Identify genes which change in either an up-down or down-up pattern, while controlling for leaf variability
	Reduced	~ Disc Number	

Table 5: DESeq2 Experimental Design Formulas for Vitis vinifera, all of which control for leaf variability via the term Disk Number.

P-values generated by DESeq2 were adjusted for multiple testing using the Benjamini-Hochberg procedure (Benjamini & Hochberg, 1995). An adjusted p-value ≤ 0.1 indicated that a gene was statistically significant. The significant genes from the union of the three designs were combined and used for further analysis. Most of the tools used for further analysis of this gene set required that the genes be labeled using gene names from the V1 annotation. The correspondence table available at

https://urgi.versailles.inra.fr/content/download/5723/43038/file/list_genes_vitis_correspondencesV3_1.xlsx

was used to convert the Vcost.v3 gene names to V1 gene names. For the instances when a V1 gene name was not available for a gene, the Vcost.v3 gene name was left unchanged. The columns for the output table of significant genes were the Vcost.v3 gene name, the V1 gene name, the 12hpc vs UTC $\log_2$ fold-change, the 24hpc vs UTC $\log_2$ fold-change, the 36hpc vs UTC $\log_2$ fold-change, and the 48hpc vs UTC $\log_2$ fold-change.

A single design formula (model) was used for the *A. thaliana* dataset (Table 6).

#	Design Formula	Explanation
1	~ Experiment + Genotype + Treatment	Identify genes while controlling for treatment, genotype, or experiment

Table 6: DESeq2 Experimental Design Formula for Arabidopsis thaliana

The "Experiment" term in the model was meant to control for variability between the 3 different experiments. The "Genotype" term was meant to control for variability between the two genotypes, Col-0 and PEN1. The "Treatment" term represented the five treatments, 48hpc,

36hpc, 24hpc, 12hpc, and UTC. In DESeq2, a Wald test was performed to generate pair-wise comparisons between a reference level and the remaining levels for each term. An example of a pair-wise comparison for "Treatment" would be 12hpc vs UTC (reference level). As with the *V. vinifera* analysis, an adjusted p-value ≤ 0.1 was used to indicated statistically significance for a gene. The log-fold changes for the pair-wise comparisons were combined into a single table where the columns represented the gene name, 12hpc vs UTC, 24hpc vs UTC, 36hpc vs UTC, and 48hpc vs UTC. The initial expression plots (see Results section) showed genes with outlier expression patterns. These outlier expression patterns could be caused by samples with extreme counts relative to other samples which raised the mean count for the respective treatment group. To overcome this issue, DESeq2 provided a method to shrink the $\log_2$ fold-change (LFC) for a gene towards zero if the information about that gene was low. In this context low information refers to having samples with low counts or having highly dispersed sample counts. This shrinkage method was applied using the *lfcShrink* function with the "type" parameter set to "Normal", which is the original DESeq2 shrinkage estimator (Love, Huber, & Anders, 2014). The shrunken LFC were combined into a single table as before and used for further analysis.

For the genotype comparison, a list of 16 genes with a documented association to the PEN1 gene was compiled (Table 7). This list was complied using the associated loci on the TAIR description for the AT3G11820 gene locus (Swarbreck et al., 2008).

Araport11 ID	Gene Name	Description
AT3G11820	*PEN1, SYR1, SYP121, ATSYR1*	Encodes a syntaxin localized at the plasma membrane (SYR1, Syntaxin Related Protein 1, also known as SYP121, PENETRATION1/PEN1). SYR1/PEN1 is a member of the SNARE superfamily proteins. SNARE proteins are involved in cell signaling, vesicle traffic, growth and development.
AT1G12360	*KEU, KEULE, SEC11*	Encodes a Sec1 protein and expressed throughout the plant. physically interacts with Syntaxin1 and is required for cytokinesis
AT1G04750	*VAMP721, VAMP7B*	Encodes vesicle-associated membrane protein 7B (VAMP7B, or VAMP721). Required for cell plate formation.VAMP721 interacts with KAT1 and KC1 K+ channels, affects channel gating and suppresses the K+ current within the physiological voltage range. Post-transcriptionally regulated by CRT1/2 under ER stress.
AT3G52400	*SYP122*	Syntaxin protein, involved in the negative regulation of defense pathways such as programmed cell death, salicylic acid signalling pathway, jasmonic acid signalling pathway
AT3G07040	*RPM1, RPS3*	Contains an N-terminal tripartite nucleotide binding site and a C-terminal tandem array of leucine-rich repeats. Confers resistance to Pseudomonas syringae strains that carry the avirulence genes avrB and avrRpm1.
AT3G48090	*ATEDS1, EDS1*	Component of R gene-mediated disease resistance in Arabidopsis thaliana with homology to eukaryotic lipases.
AT4G16144	*AMSH3*	Encodes AMSH3, a deubiquitinating enzyme that hydrolyzes K48- and K63-linked ubiquitin chains in vitro. Required for intracellular trafficking and vacuole biogenesis.
AT4G26090	*RPS2*	Encodes a plasma membrane protein with leucine-rich repeat, leucine zipper, and P loop domains that confers resistance to Pseudomonas syringae infection by interacting with the avirulence gene avrRpt2. RPS2 protein interacts directly with plasma membrane associated protein RIN4 and this interaction is disrupted by avrRpt2. The mRNA is cell-to-cell mobile.
AT3G08530	*ATCHC2, CHC2*	CHC2 heavy chain subunit of clathrin. Involved in vesicle mediated trafficking. Mutants show reduced rates of endocytosis and defects clathrin mediated exocytosis Mutants have increased drought tolerance due to defects in stomatal movement.
AT3G11130	*ATCHC1, CHC1*	CHC1 heavy chain subunit of clathrin. Involved in vesicle mediated trafficking. Mutants show reduced rates of endocytosis and defects clathrin mediated exocytosis. Mutants also have increased dehydration tolerance which may be related to the overall slower stomatal aperture dynamics. Overall growth is affected.
AT1G59870	*ABCG36, PDR8, PEN3*	ATP binding cassette transporter. Localized to the plasma membrane in uninfected cells. In infected leaves, the protein concentrated at infection sites. Contributes to nonhost resistance to inappropriate pathogens that enter by direct penetration in a salicylic acid–dependent manner. Required for mlo resistance. Has Cd transporter activity (Cd2+ extrusion pump) and contributes to heavy metal resistance. The mRNA is cell-to-cell mobile.
AT5G61210	*ATSNAP33, SNAP33, SNP33*	Membrane localized t-SNARE SNAP25 homologue, probably involved in cytokinesis and cell plate formation The mRNA is cell-to-cell mobile.
AT2G44490	*BGLU26, PEN2*	Encodes a glycosyl hydrolase that localizes to peroxisomes and acts as a component of an inducible preinvasion resistance mechanism. Required for mlo resistance. The mRNA is cell-to-cell mobile.
AT2G20990	*SYT1, SYTA*	Encodes a protein specifically localized to the ER-PM boundary with similarity to synaptotagmins, a class of membrane trafficking proteins.
AT5G46240	*ATKAT1, KAT1*	Encodes a potassium channel protein (KAT1). ABA triggers KAT1 endocytosis both in epidermal cells as well as guard cells. Upon removal of ABA, KAT1 is recycled back to the plasma membrane.
AT2G33110	*VAMP723*	Member of VAMP72 Gene Family
AT4G32650	*KAT3, KC1*	Encodes KAT3, a member of the Shaker family of voltage-gated potassium channel subunits. Does not form functional potassium channel on its own. Involved in down-regulating AKT1 and KAT1 channel activity by forming heteromers with AKT1 or KAT1.
AT4G15340	*04C11, ATPEN1, PEN1*	Encodes a protein that catalyzes the production of the tricyclic triterpene arabidiol when expressed in yeast.

Table 7: Genes with Documented Association to PEN1 gene

For both hosts, R v.3.5.2 was used for the analysis (R Core Team, 2018). The ggplot2 and pheatmap packages were used to produce the plots and heatmaps representing the gene expression levels across the treatment groups (H. Wickham, 2016), (Kolde, 2019).

Outlier analysis

During the gene counting step, outlier read counts can occur due to a number of factors, including RNA extraction methods, experimental design, and the specified settings for the counting software. The presence of outliers in an RNA-Seq experiment can drastically

influence differential expression testing results (Li & Tibshirani, 2013). To screen for outliers, DESeq2 utilizes the Cook's distance to measure how a single observation impacts the count prediction model. This method of screening is generally sufficient, but can leave DESeq2 potentially susceptible to extreme outliers present in skewed or biased datasets (George, Bowyer, Crabtree, & Chang, 2015). To overcome this limitation, George et al. (2015) suggested implementing an iterative leave-one-out approach to outlier detection. This approach was applied to both datasets.

Due to the high number of biological replicates in the *V. vinifera* dataset, a threshold level of 90 flagged genes was used to qualify a sample for removal from the analysis. Sample count quality controls were performed using the plot functions from DESeq2 (Love, Huber, & Anders, 2014), ggplot2 (H. Wickham, 2016), and base R (R Core Team, 2018). A PCA plot was performed on the *V. vinifera* dataset, with the samples colored by experiment. Table 8 shows the number of biological replicates per treatment for the *V. vinifera* dataset, after removing the outlier samples. The differential expression analysis was performed using DESeq2 with the outliers removed.

Treatment	UTC	12hpc	24hpc	36hpc	48hpc
# of Samples	10	10	10	9	10

Table 8: Vitis vinifera samples used for DESeq2 analysis, after filtering samples for outlier counts

For the *A. thaliana* dataset, the iterative leave-one-out approach (George et al., 2015) was repeated. Due to the lower number of biological replicates in this dataset, all samples were included in the analysis. The default settings for DESeq2 were used for outlier identification and correction. Table 9 shows the number of biological replicates for the *A. thaliana* dataset used for the analysis.

Treatment	UTC	12hpc	24hpc	36hpc	48hpc
# of Col-0 Samples	9	9	9	9	9
# of Pen-1 Samples	9	9	9	9	9

Table 9: Arabidopsis thaliana samples used for DESeq2 analysis, with no outlier count filtering

Cluster Analysis

A cluster analysis was performed to determine groups of genes that had similar expression patterns across treatments. The R package "Mfuzz" (Futschik, 2005) provided a function, kmeans2, which performed a k-means clustering based on the number of centroids provided. A workflow was devised for objectively choosing the optimum number of centroids and well as the clusters of interest. The workflow evaluated using between 6 and 15 cluster centroids, inclusively.

Objective Cluster Analysis Workflow:

1) For each cluster centroid setting (6:15):
   A. For each time point, in each cluster:
      1) Calculate the mean LFC for that time point.
      2) Calculate the LFC deviation for that time point.
   B. Calculate the mean (average) deviation across all the time points for each cluster.
2) Using the mean deviation for each cluster, each cluster centroid setting was ranked by its median mean deviation. The minimum mean deviation for this ranking represented the optimum centroid setting because all of the clusters are group more tightly together compared to the other settings.
3) Using the optimum cluster centroid setting:
   A. Calculate the absolute mean LFC at the treatment group of interest* between the clusters.
   B. Order the clusters by decreasing absolute mean LFC.

*The treatment group of interest depended on the organism due to time of maximum resistance phenotype. For *V. vinifera* the 24hpc treatment group was used. For *A. thaliana* the 12hpc treatment groups was used.

Functional Annotation of Significant Genes

The initial proposal for this project included the Gene Ontology (GO) (The Gene Ontology Consortium, 2019) as the sole source of functional information for the significant genes between the two datasets. At the time of this project's completion, however, the functional information maintained by the GO for *V. vinifera* was limited. Of the total number of differentially expressed genes for *V. vinifera*, about 44% (1161 out of 2654 genes) were either unclassified or their 'gene id' was not found in the database. As an alternative, functional annotations for genes were determined using information from the VitisNet knowledgebase (Grimplet et al., 2012). Gene accession numbers were used as unique identifiers to map each to associated gene ontology term, plant ontology terms, pathways associations, known *Arabidopsis thaliana* homologs, and other functional information. Genes that did not map to any VitisNet information were annotated using the GO in an attempt to provide some type of functional information.

The functional information maintained by the GO for *A. thaliana* was of higher quality. Of the total number of differentially expressed genes for *A. thaliana*, only 19% (138 out of 726 genes) were either unclassified or their gene id was not found in the database. Functional terms were also more specific than those terms for *V. vinifera*. Significant genes between both the Col-0 and PEN1 genotypes were also functionally annotated to aid in the comparison.

A gene enrichment analysis was performed for both organisms using the agriGO (Tian et al., 2017) platform, a web-tool optimized for agricultural gene ontology analyses. For *V. vinifera*, the Gramene Release 50 gene list was used as the background reference gene list. For *A. thaliana*, the TAIR10 (2017) gene list was used as the background reference.

Pathway Analysis

The focus of this study was to determine the pathways affected by cold-SIDR. Pathway annotations from the PlantCyc database (Schläpfer et al., 2017) were used to determine the pathways affected in *V. vinifera* and *A. thaliana*. The databases for each organism were obtained and integrated into the web resource VitisPathways (Osier, 2016). Each cluster of significant genes was submitted separately for analysis. Enriched pathways for each cluster were determined using 1000 permutations and a permuted p-value of <0.05. The results were converted into a tab-separated value (TSV) file and read into R for further data processing. Heatmaps were made in R v.3.5.2 using the pheatmap package (R Core Team, 2018; Kolde, 2019). The color scale for the heatmaps were based on the permuted p-value.

Organism Comparison

Gene homolog information was used to assist making a comparison between the two organisms. *A. thaliana* homologs for *V. vinifera* genes were available from the VitisNet knowledgebase (Grimplet et al., 2012). For *V. vinifera* genes missing homologs information, potential homologs were retrieved from the Ensembl BioMart portal (Smedley et al., 2015) using the plants genes release #42. Still separated in clusters, gene homologs from either *V. vinifera* or *A. thaliana* were mapped to the significant genes for both datasets. The homologs genes for the clusters were submitted to VitisPathways using the appropriate pathway database. Enriched pathways were determined using 1000 permutations and a permuted p-value of <0.05. Result processing and visualization were the same as methods used for the pathways analysis for the significant genes.

Results

Preprocessing

For the *V. vinifera* dataset, the initial and final read counts following the preprocessing steps are summarized in Table 10.

	Minimum	1st Quartile	Median	3rd Quartile	Maximum	Mean	Standard Dev.
Initial Counts	2,059,331	3,837,500	5,113,182	6,755,654	9,940,126	5,321,273	1,843,200
Final Counts	1,836,703	3,435,506	4,608,237	6,096,526	8,953,438	4,789,809	1,673,346

Table 10: Vitis vinifera Preprocessing Read Count Summary Statistics per Replicate

There was a large gap between the minimum and maximum read count for the samples, which persisted after the FastX processing. Also, the mean number of reads is larger than the median both before and after processing. This is a potential indication that there are more samples with read counts below the mean than above the mean.

For the *A. thaliana* dataset, the initial and final read counts following the preprocessing steps are summarized in Table 11.

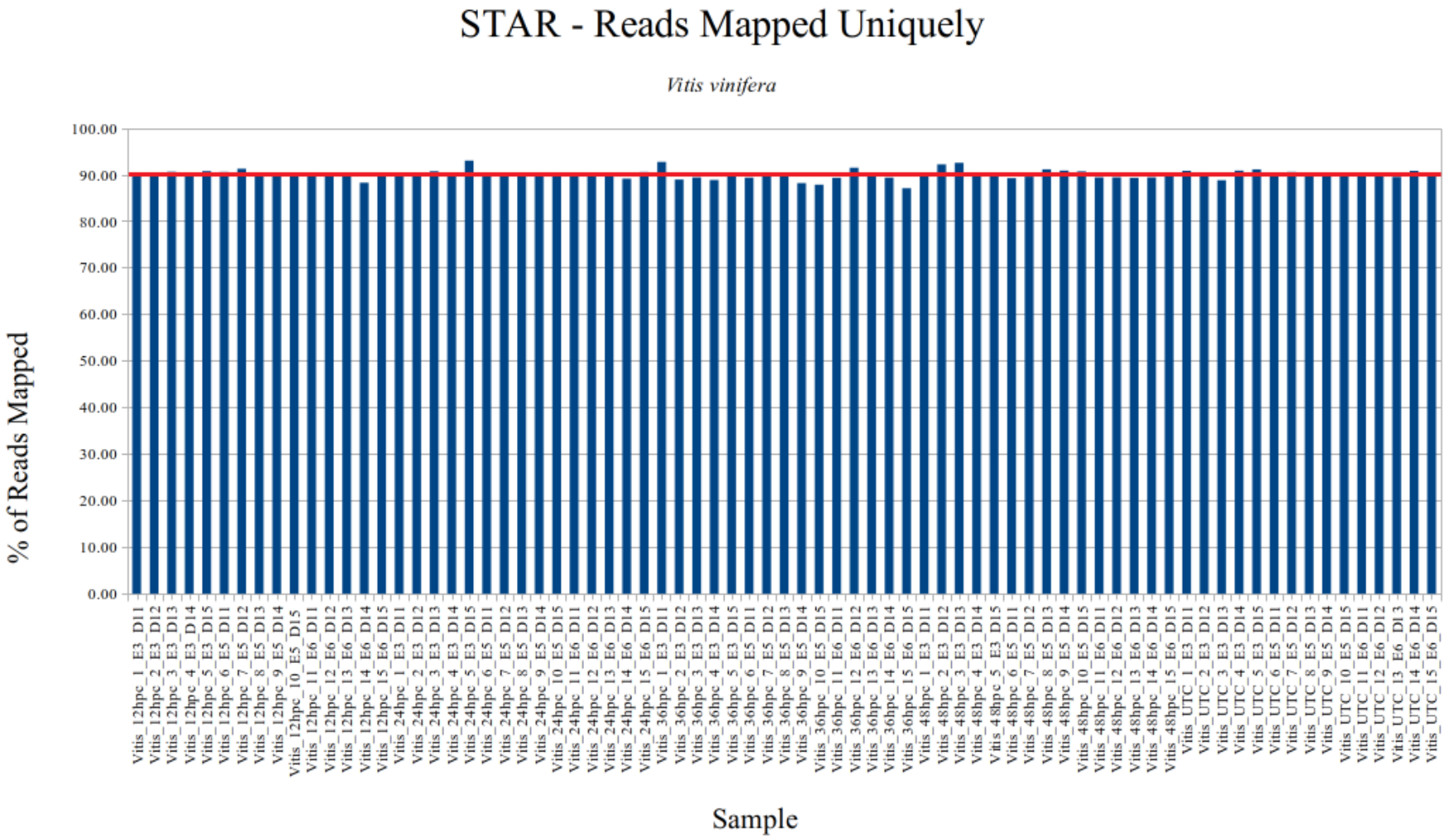
	Minimum	1st Quartile	Median	3rd Quartile	Maximum	Mean	Standard Dev.
Initial Counts	1,255,504	2,593,330	3,099,480	3,675,016	7,218,345	3,310,981	1,123,431
Final Counts	1,083,552	2,338,518	2,680,044	3,293,006	6,481,624	2,948,336	1,006,844

Table 11: Arabidopsis thaliana Preprocessing Read Count Summary Statistics

Similar to the *V. vinifera* dataset, there is a large gap between the maximum and minimum sample read counts which remains after the FastX processing. Again, similar to *V. vinifera*, the read counts has a distribution that is to the right. The standard deviation is smaller than that of the *V. vinifera* dataset indicating that the read counts for the *A. thaliana* samples cluster more closely to the mean count. This is also likely due to the low read counts across the samples in the *A. thaliana* dataset.

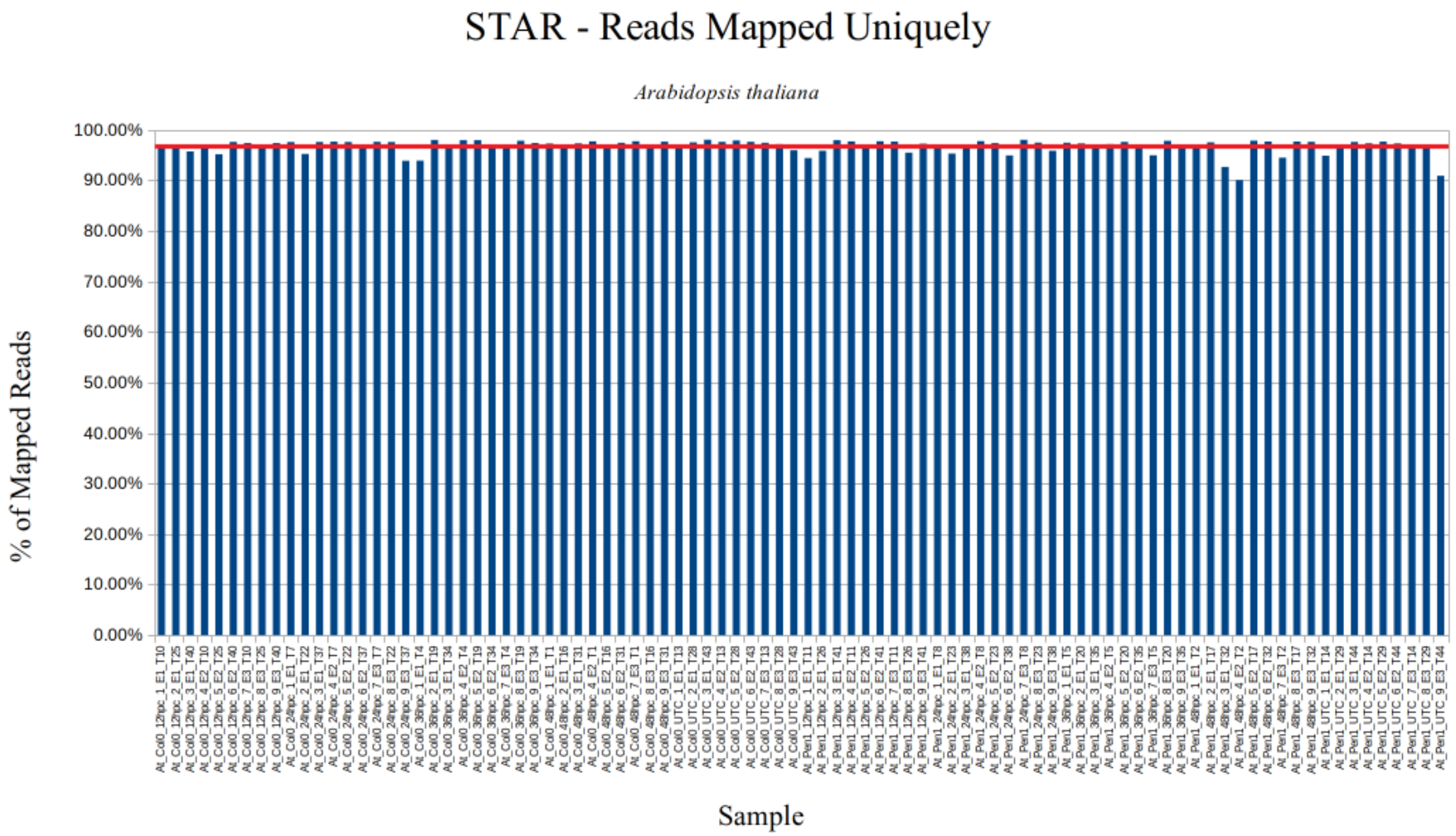
Alignment

For all 90 *V. vinifera* samples, the STAR software produced a high number of read alignments, with 90.1%, of the reads on average uniquely mapping to the genome (Figure 10). The minimum value was 87.1%. The percentages of uniquely mapped reads are shown. The mean percent of uniquely mapped reads for the *A. thaliana* dataset was 96.8% (Figure 11), with a minimum of 90.1%. For both datasets, the minimum percent of mapped reads were within the range of the expected mapped reads recommended by Conesa et al. (2016) to indicate high overall sequence accuracy (70-90%).



** Red Line Indicates the Mean Percent of Uniquely Mapped Reads*

Figure 10: STAR Reads Uniquely Mapped to Vitis vinifera



** Red Line Indicates the Mean Percent of Uniquely Mapped Reads*

Figure 11: STAR Reads Uniquely Mapped to Arabidopsis thaliana

Gene Counting

For the *V. vinifera* dataset, Figure 12* shows the distribution of the reads counts per sample. The samples had a mean of 3,059,445 reads counted and assigned to distinct gene positions (Figure 12-green line). The smallest number of reads counted was 1,100,654 and the largest was 5,896,608. The mean number of reads categorized as uncounted was 1,664,959 (Figure 12-blue line). The maximum number of uncounted reads for a sample was 3,352,702, though this sample also had the largest number of total reads.

***Note**: Figures 12-14 show the distribution of the reads counts per sample show the total reads, the total reads counted, and the total uncounted reads. These regions are organized from front to back, with the total uncounted in front, total counted behind that, and the total number of reads in the back. If the region of uncounted reads is placed at the end of the top region of the counted reads, the total of the two will equal the total number of reads for the sample.

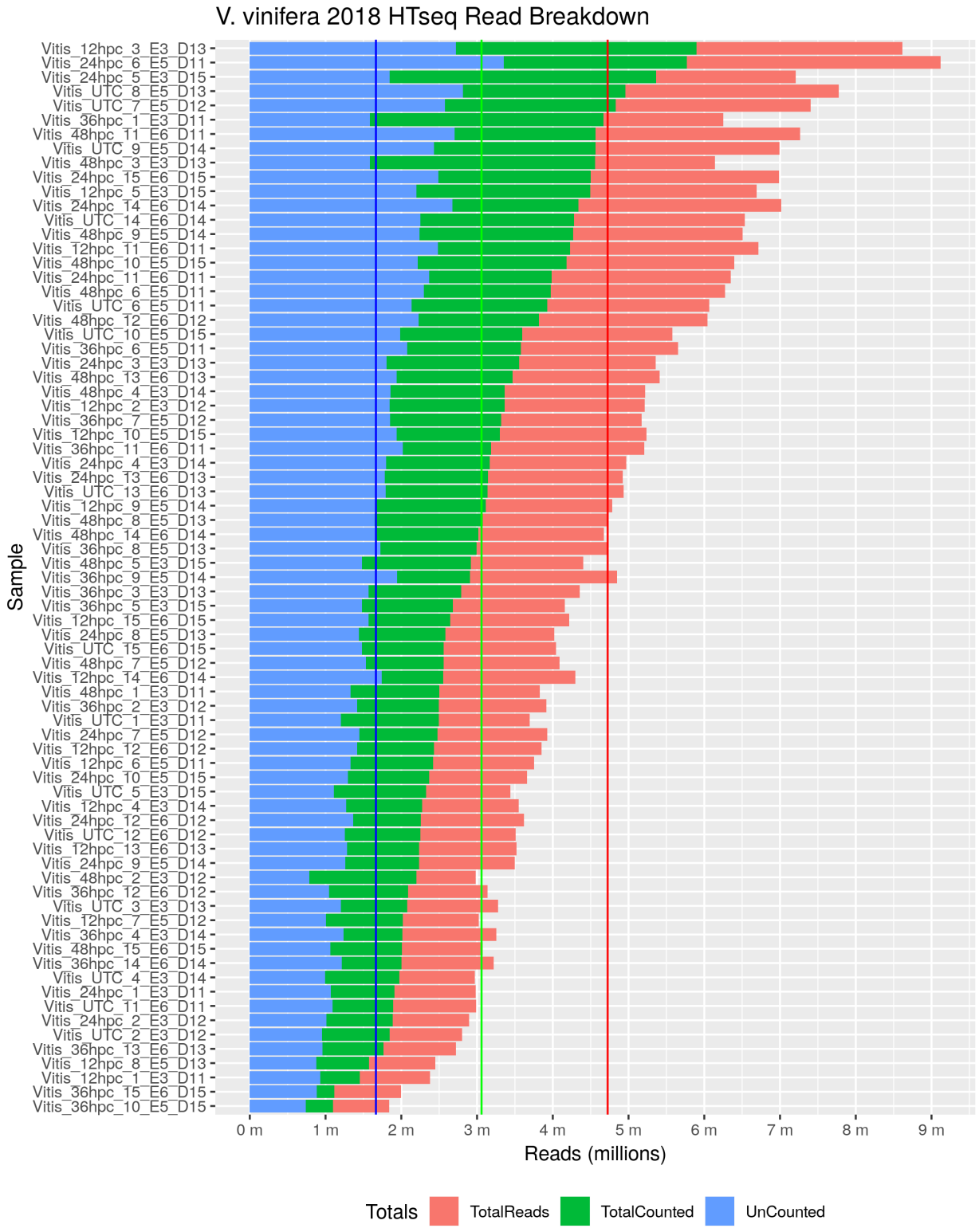


Figure 12: Vitis vinifera - HTSeq Read Count Distribution

For the Col-0 samples, Figure 13* shows the distribution of reads per sample, colored by the total number of read, total reads counted, and total of reads uncounted. The samples had a mean of 3,059,445 reads counted and assigned to distinct gene positions (Figure 13-green line). The smallest number of read counted was 1,100,654 and the largest was 5,896,608. The mean number of reads uncounted was 312,498 reads (Figure 13–blue line).

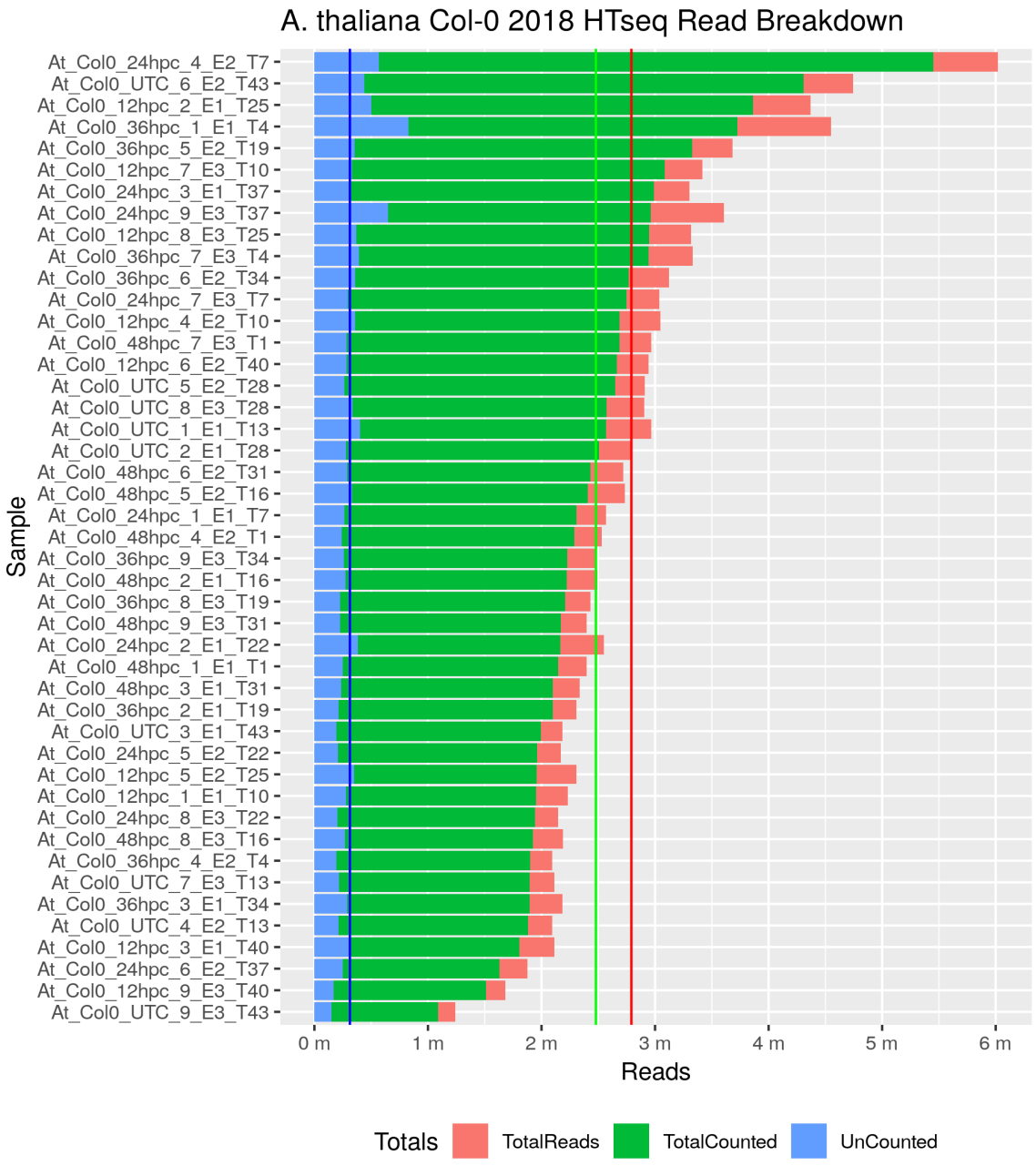


Figure 13: Arabidopsis thaliana - Col-0 - HTSeq Read Count Distribution

Figure 14* shows the distribution of reads per sample for the PEN1 samples. The PEN1 samples had a mean of 2,783,801 of reads counted (Figure 14-green line). For the PEN1 samples, the minimum number of reads counted was 959,305 and the maximum number was 5,608,963. The mean number of reads uncounted was 381,325 reads (Figure 14–blue line).

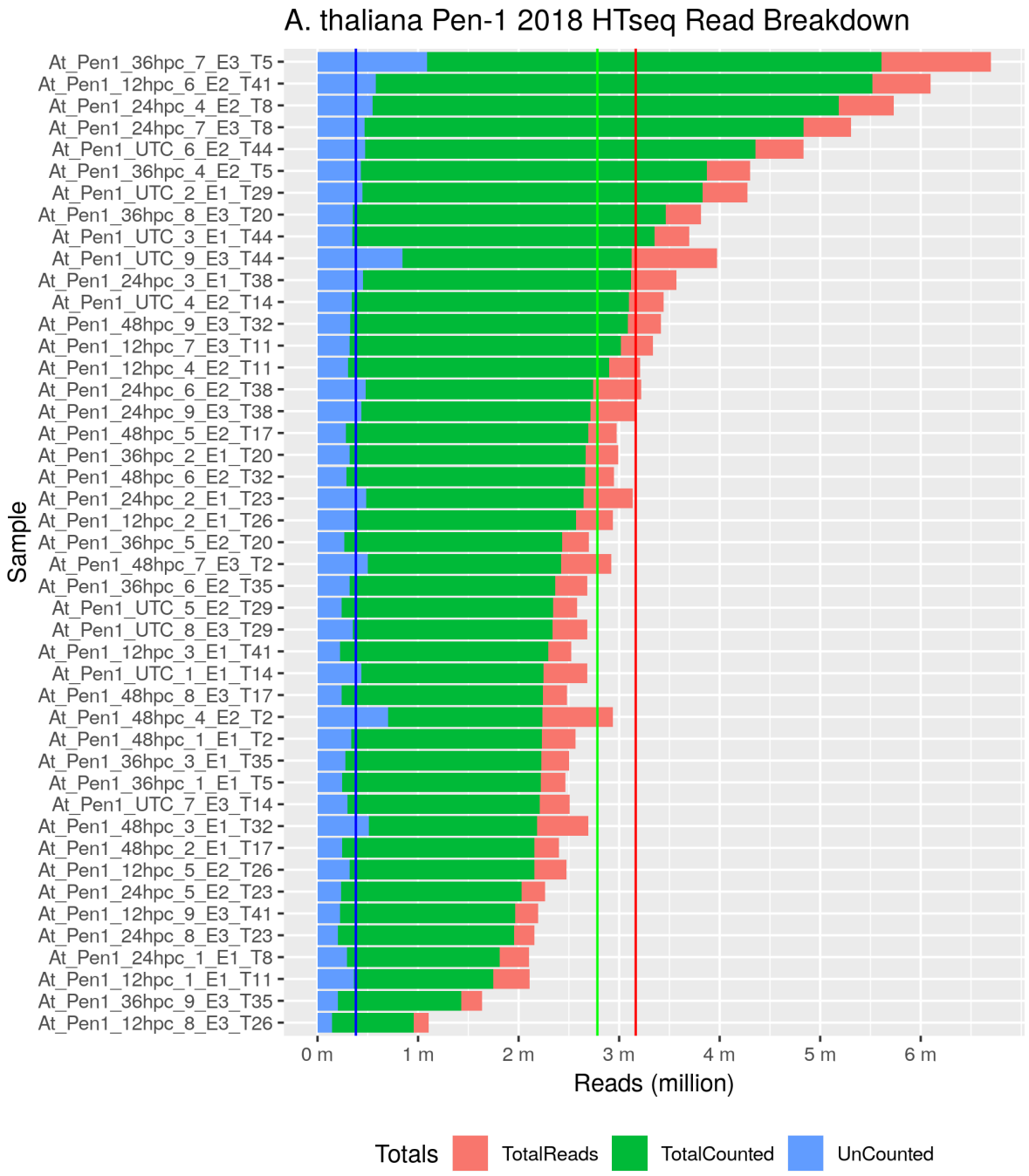


Figure 14: Arabidopsis thaliana - PEN1 - HTSeq Read Count Distribution

A pattern that could be observed between the two datasets was that more total reads generally led to more reads being counted for a gene. It can also be observed that the *V. vinifera* has more reads that are uncounted compared to either the Col-0 or PEN1 datasets. Reads that are classified as "uncounted" either map to a region of the genome that does not contain an annotated gene, or that the read mapped to multiple regions or multiple genes.

Differential Expression

Prior to the differential expression analysis, a PCA plot was performed on the *V. vinifera* dataset, with the samples colored by experiment (Figure 15). The samples from experiment #3 were clustered away from the other two experiments. Based on this clustering, samples from experiment #3 were removed.

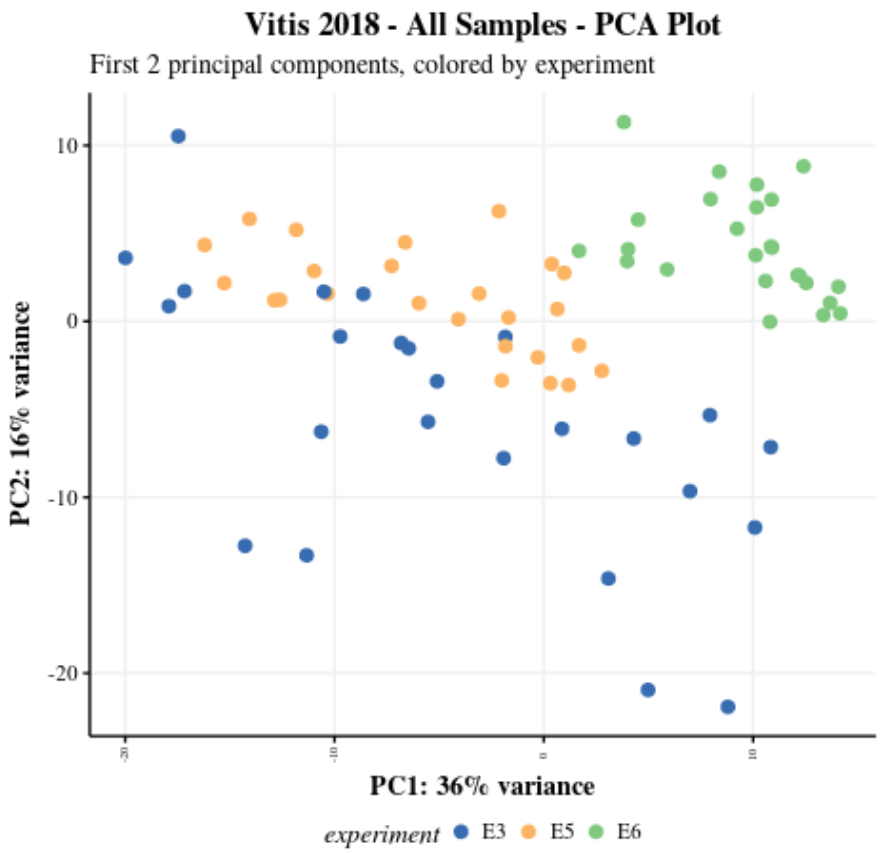


Figure 15: Vitis vinifera PCA for all samples

For the *V. vinifera* dataset, the 3 model approach as described in Methods identified 2654 significant genes, of which 55 genes were identified in all three models. The exact number of genes generated by each model is represented in Figure 16.

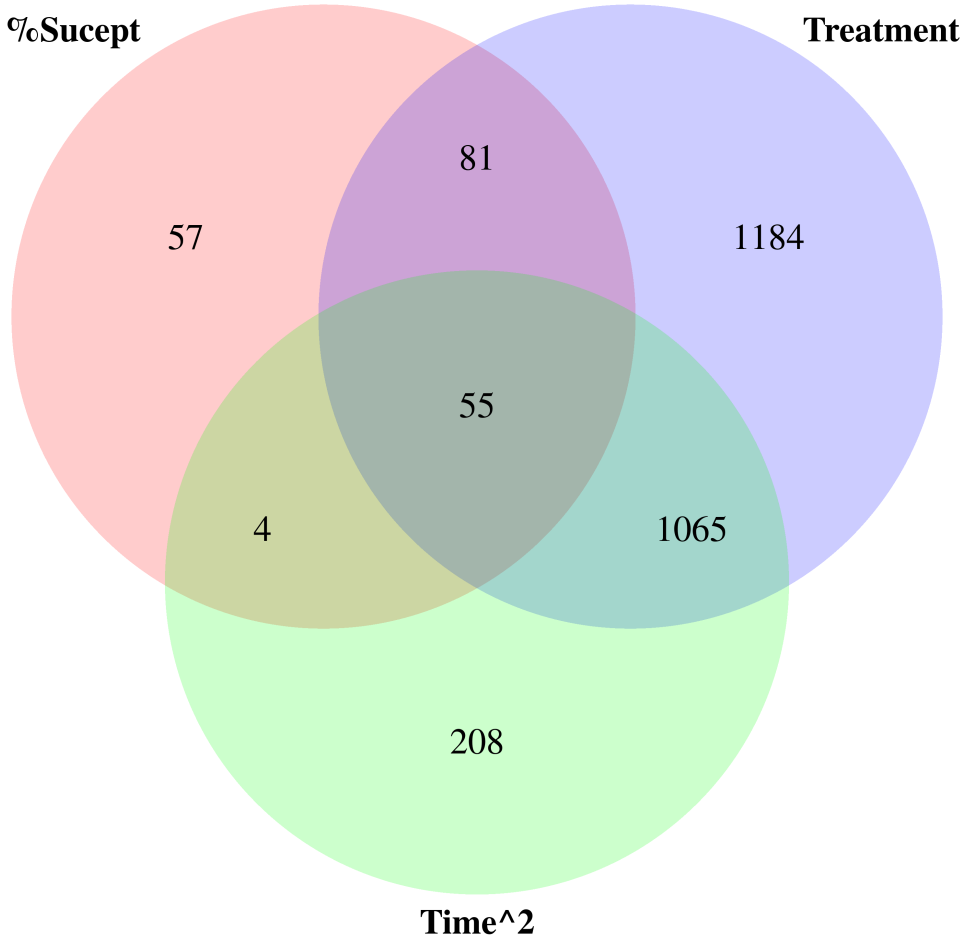


Figure 16: Venn diagram showing overlapping significant genes by model

The "treatment" model produced the largest number of differentially expressed genes, followed by the "time^2" model, and then the percent susceptibility model. Due to the nature of the likelihood ratio test, the exact number of differentially expressed genes at each treatment cannot be determined. Figure 17 shows the expression profiles for the DEGs relative to the UTC treatment group. The gene expressions were converted to $\log_2$ fold-change (LFC) to make the results more intuitive. This way, the expression ratios are treated symmetrically, so a gene that is

up-regulated by a factor of 2 has a $\log_2$ (ratio) of 1 and a gene that is down-regulated by a factor of 2 has a $\log_2$ (ratio) of -1. A majority of the DEGs had a LFC between -1 and 1. Some genes extended beyond this threshold, reaching a maximum LFC of 4.11 or a minimum LFC of -2.42.

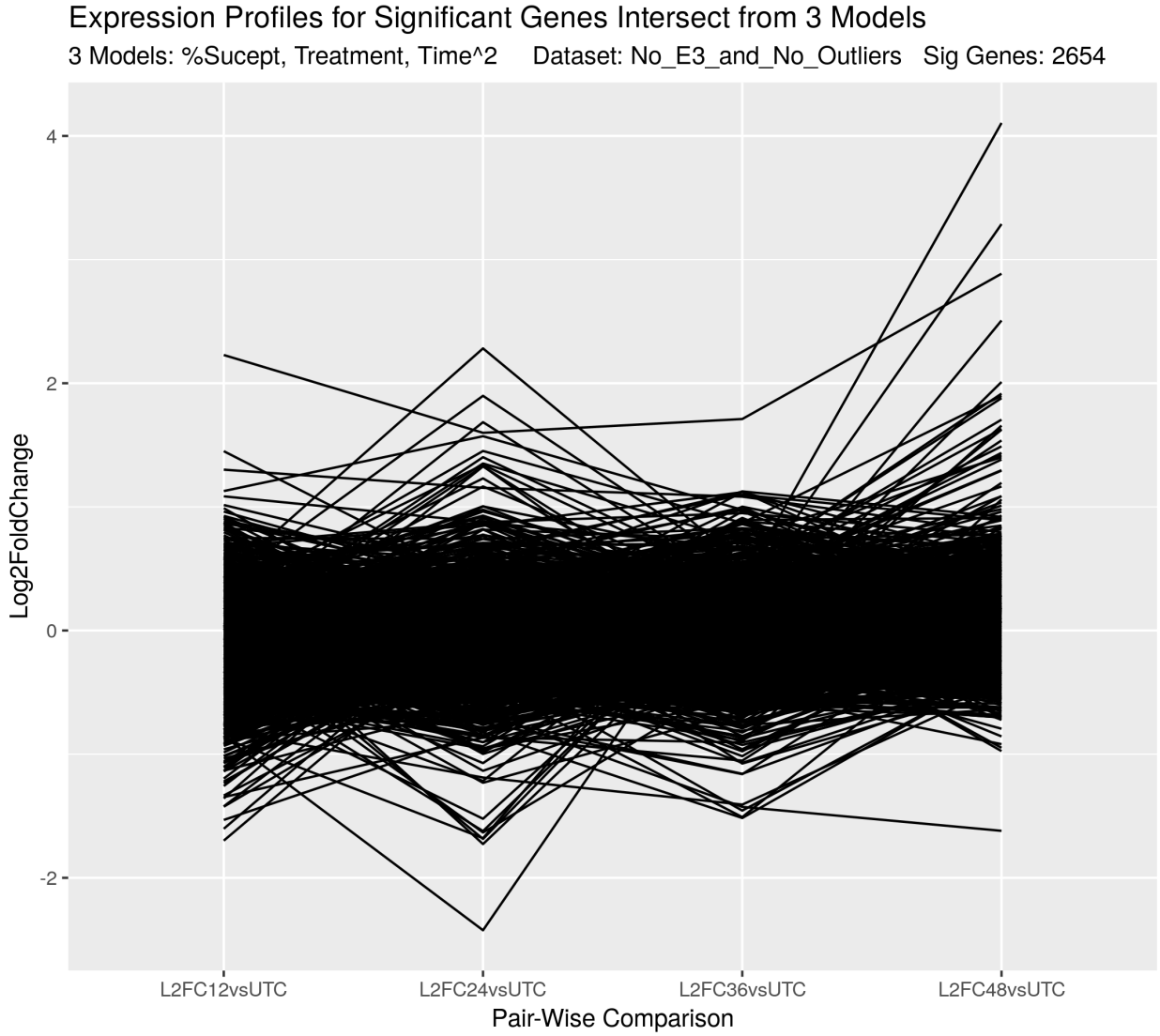


Figure 17: Vitis vinifera Expression Profiles for Significant Genes

Genes with the largest positive LFC occurred in the 48hpc treatment group. Genes with the largest negative LFC occurred in the 24hpc treatment group. There were groups of genes that appeared to follow similar expression patterns of up or down regulation. The heatmap

shown in Figure 18 illustrates this, with the colors indicating the log_2-fold change relative to the UTC. Finding these groups of similar expressed genes were the focus of the cluster analysis.

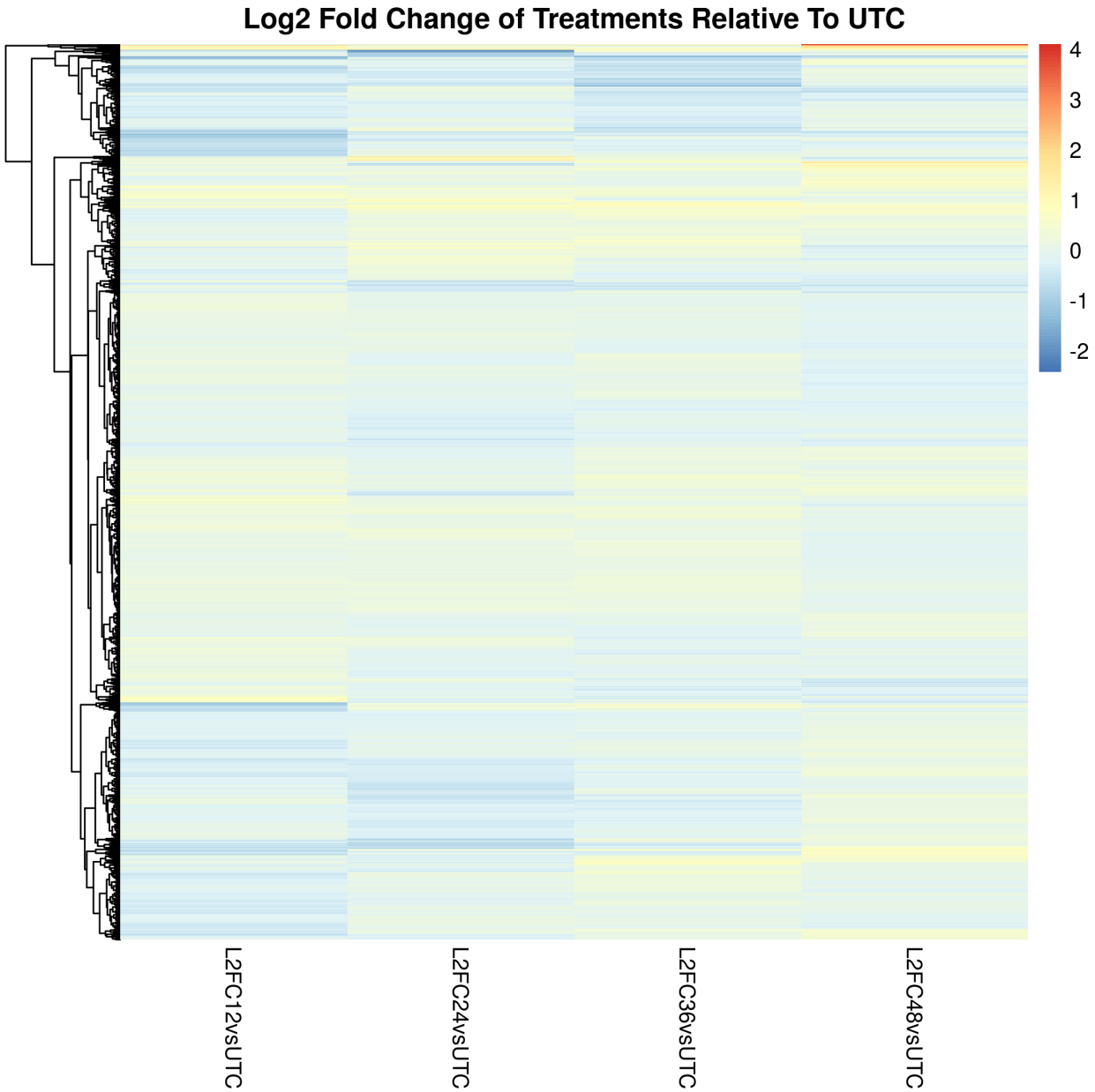


Figure 18: Heatmap showing all Significant Vitis vinifera genes, colored by log_2 fold-change

For the *A. thaliana* dataset, the Wald Analysis identified 731 significant differentially expressed genes. The Wald test produces a count of genes that were differentially expressed for each pair-wise comparison. For the different treatments, the 24hpc versus UTC pair-wise

comparison produced the largest number of differentially expressed genes, with 640. The 12hpc versus UTC pair-wise comparison produced the second largest number of genes with 111. The 48hpc versus UTC pair-wise comparison identified 5 genes as differentially expressed. No genes were reported as differentially expressed for the 36hpc treatment relative to the UTC. The unshrunken expression patterns for these genes can be seen in Figure 19.

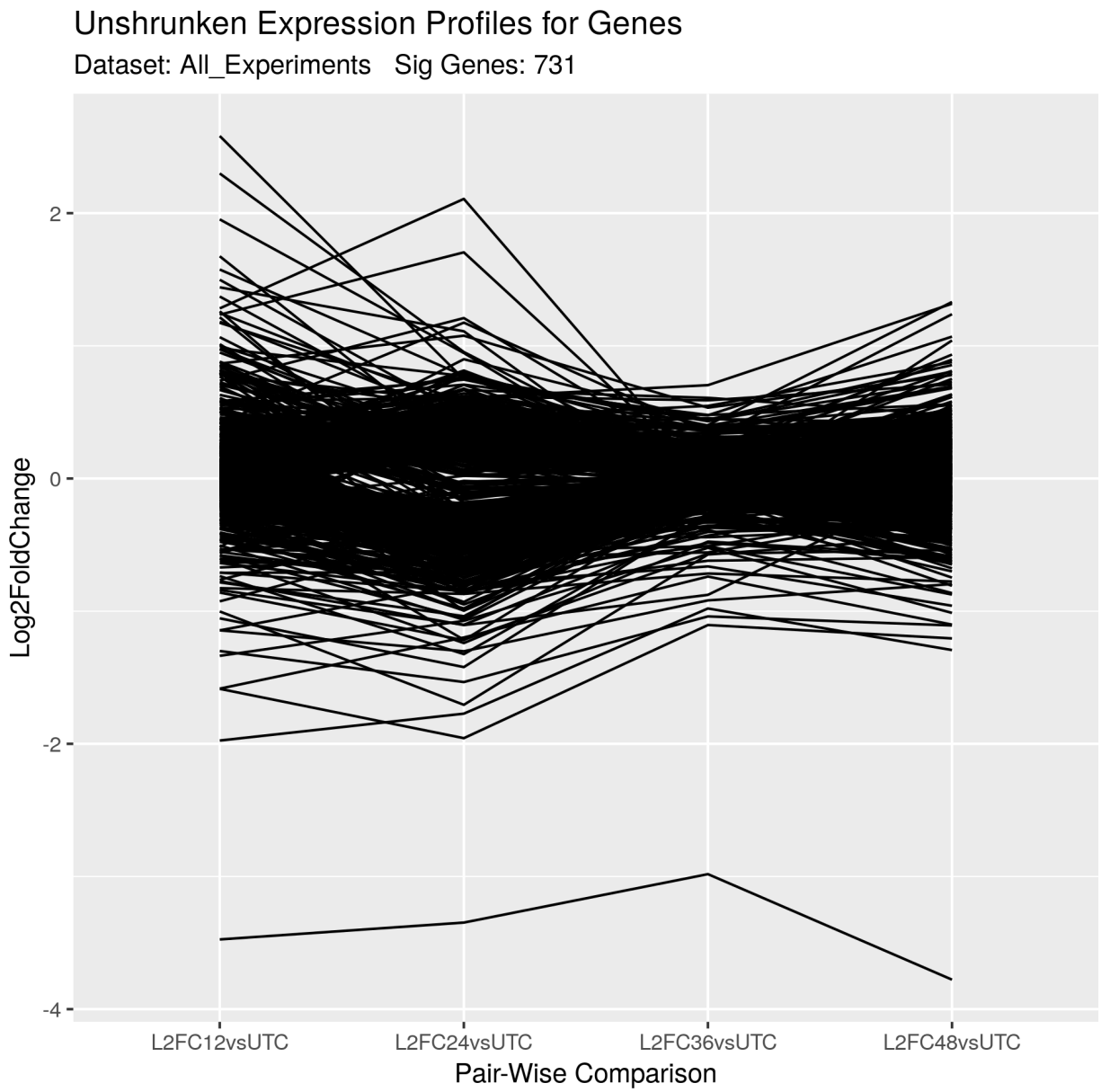


Figure 19: Arabidopsis thaliana Unshrunken Expression Profiles for Significant Genes

A single gene (AT4G12500) shows a large down-regulation in all the treatments relative to the UTC. Further investigation showed that the gene had a single sample in the UTC group with a gene count above 600, while the gene count for the other samples, including the other UTC samples were below 50. This outlier sample can be seen in Figure 20.

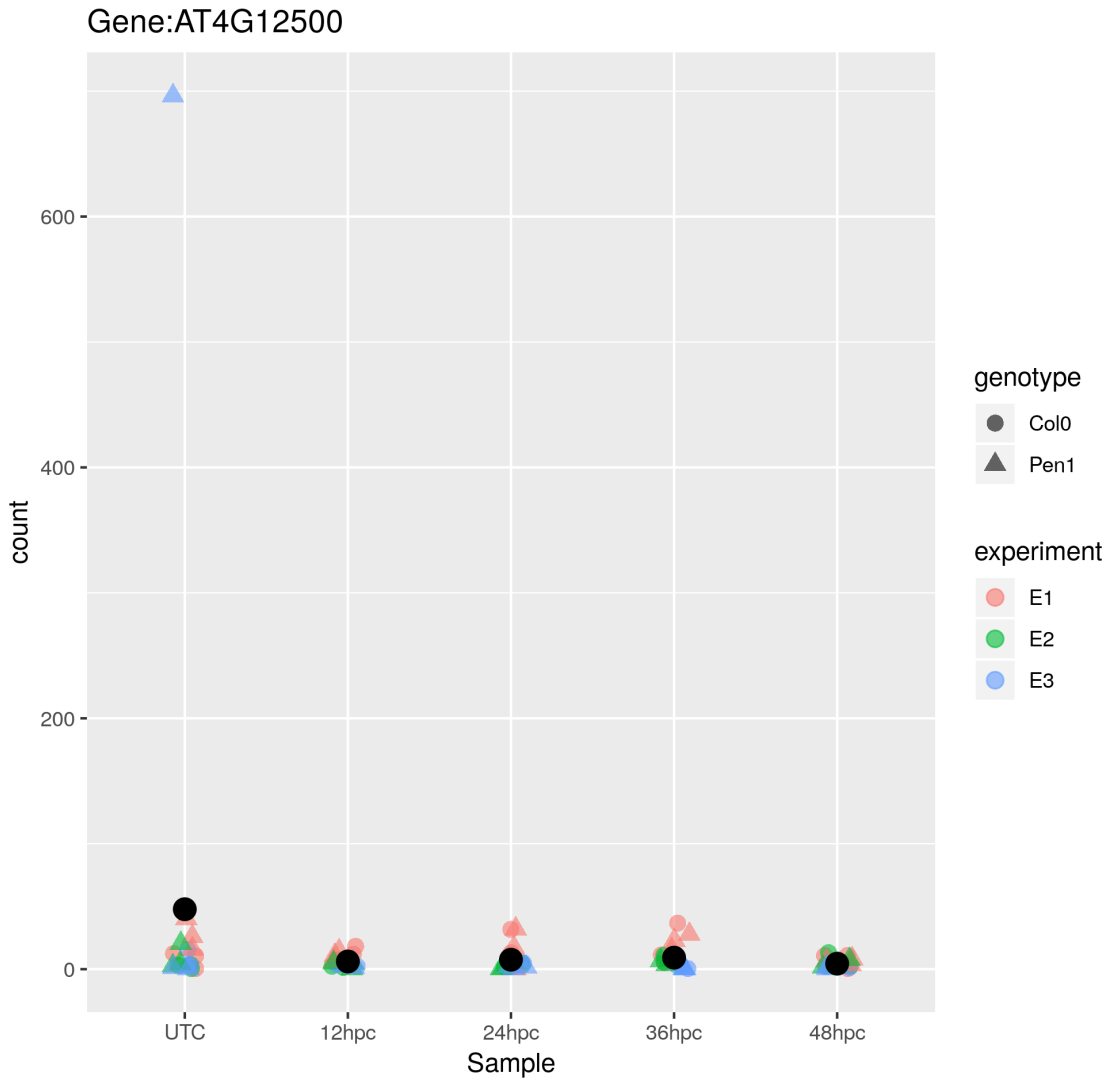


Figure 20: Arabidopsis thaliana - Gene Count Outlier

Other genes were found to have outlier samples either inflating or deflating the gene's expression profile. As described in Methods and Materials, the LFCs for the gene expression profiles were shrunken based the evidence for the gene.

Figure 21 shows the resulting shrunken expression profiles for the *A. thaliana* dataset. The expression profiles have a more compact LFC than that of the *V. vinifera* results. The maximum LFC was 1.00 and the minimum LFC was -0.57. The shape is similar to the expression profiles for the *V. vinifera* with genes being up or down regulated at 24hpc and transitioning to an expression pattern similar to the UTC at 36hpc. In contrast to the *V. vinifera*, the genes with the largest LFC reside in the 12hpc treatment group.

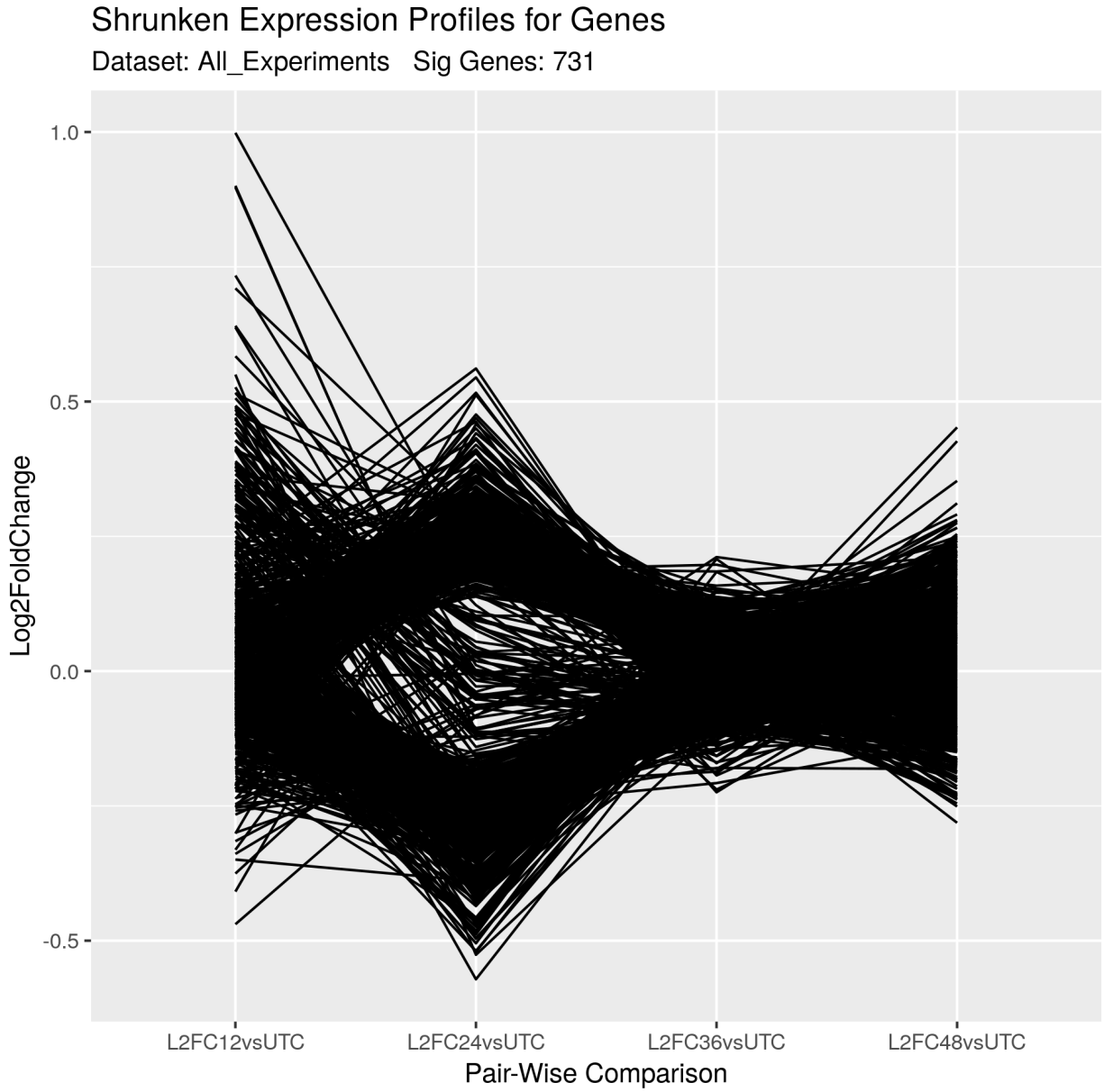


Figure 21: Arabidopsis thaliana Shrunken Expression Profiles for Significant Genes

The pair-wise comparison between the Col-0 and PEN1 genotypes produced 134 significant genes that were differentially expressed. The LFC for the significant genes were also shrunken to reduce the effect of outlier samples. The minimum shrunken LFC was -0.78 and the maximum was 0.72. The PEN1 gene had the largest negative LFC relative to the Col-0 genotype. The sixteen genes with a documented association with PEN1 (Table 7) were not identified as significantly differentially expressed relative to the Col-0 genotype.

Cluster Analysis

The objective cluster analysis performed on the gene expression values produced an optimum number of 13 clusters for the *V. vinifera* dataset (Figure 22). Taking into account previous observation of the cold-SIDR phenotypic response, the clusters of interest were determined by focusing on genes with expression changes at the 24hpc treatment. Six of the 13 clusters had negative mean LFC at 24hpc. The remaining 7 clusters had a positive mean LFC at 24hpc. The 13 clusters were ranked based on their absolute-LFC at 24hpc. Cluster #2 showed the largest absolute-LFC with mean LFC of -0.98. The general pattern for cluster #2 is that the genes showed a down-regulation at 24hpc with relatively low differential expression at the 12hpc, 36hpc, and 48hpc treatments. Cluster #3 showed the second largest absolute-LFC with a mean LFC of 0.75. The general pattern for cluster #3 is that gene showed up-regulation at 24hpc with relatively low differential expression in the other treatment groups. Cluster #13 showed the third largest absolute-LFC with a mean LFC of 0.41. Cluster #13 also included the genes with the largest LFC relative to the UTC so those genes were investigated more closely as well.

For the *A. thaliana*, the objective cluster analysis also determined an optimum number of 13 clusters (Figure 23). Previous observations (Weldon et al., manuscript in preparation) of the cold-SIDR phenotype in *A. thaliana* suggested the clusters of interest should have expression

changes during the 12hpc or 24hpc treatment, not at the 36 or 48hpc treatment. Some clusters showed up-regulation in 12hpc and down-regulation in 24hpc, with minimal expression in either 36hpc or 48hpc. Genes in clusters #2, #4, and #6 showed up-regulation in the 12hpc treatment, while genes in cluster #9 showed down-regulation at 12hpc. In the 24hpc treatment group, clusters #4, #9, and #10 showed up-regulation, while clusters #6, #3, and #11 showed down-regulation.

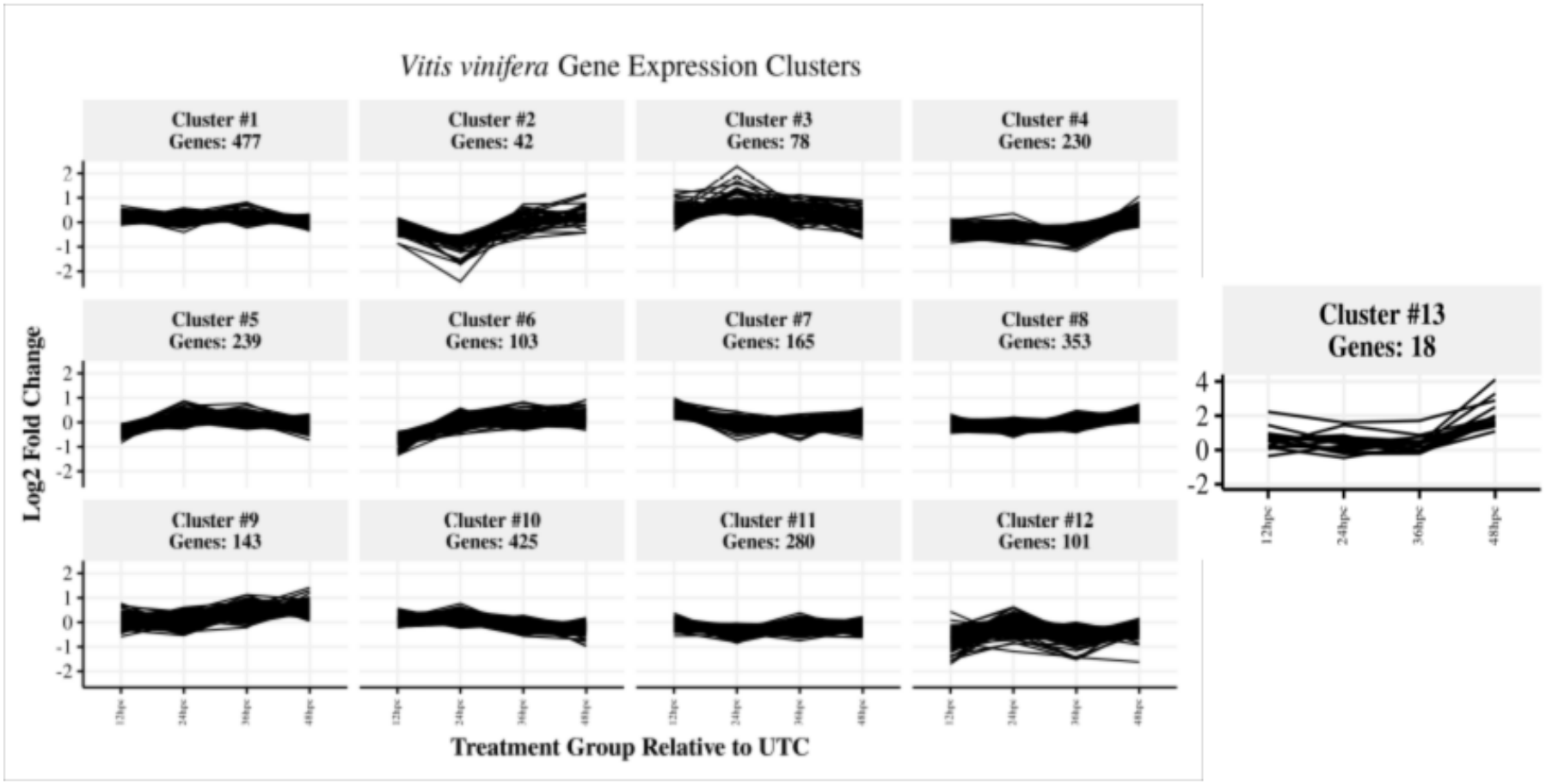


Figure 22: Vitis vinifera Gene Expression Clusters

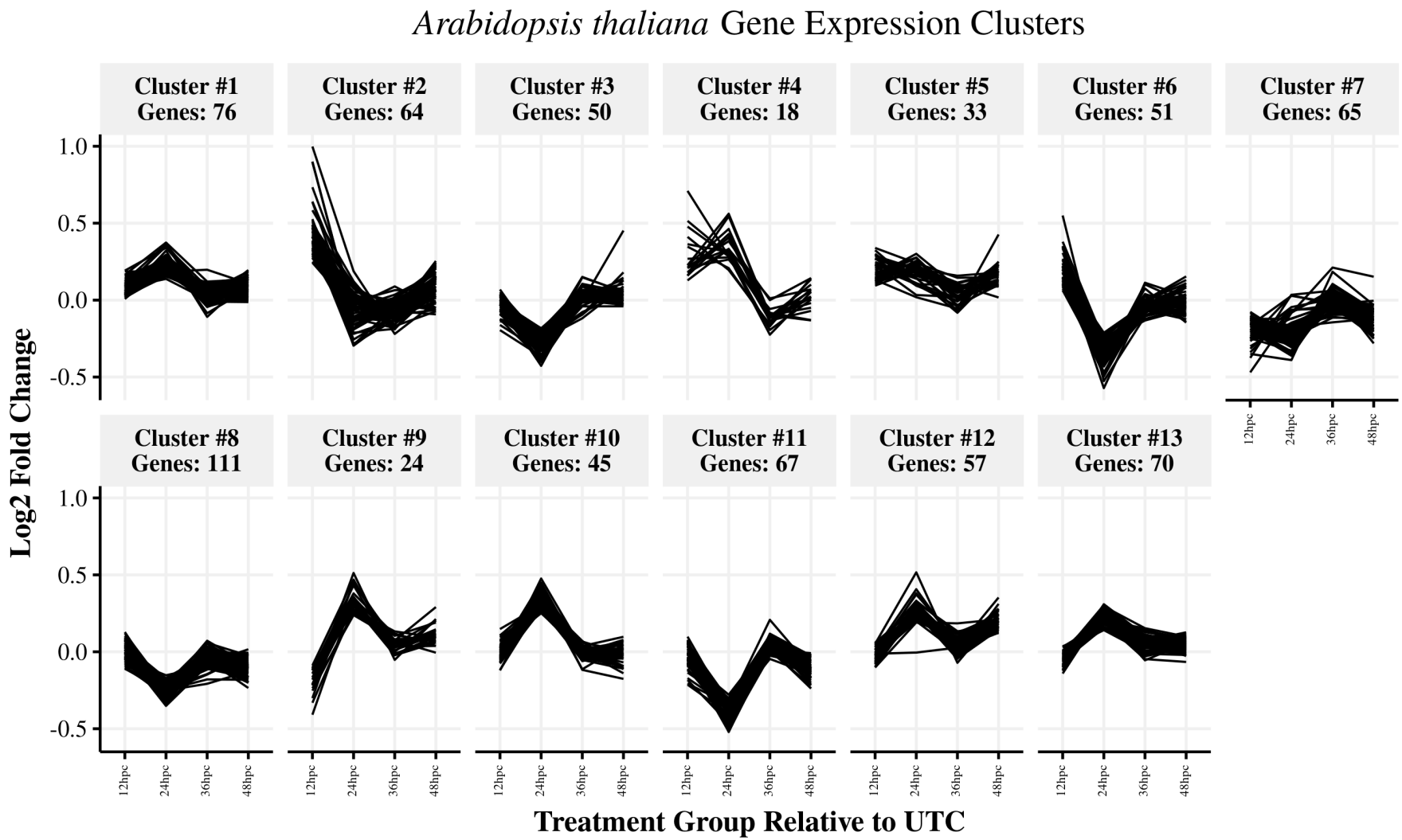


Figure 23: Arabidopsis thaliana Gene Expression Clusters

Functional Annotation

The significant *V. vinifera* genes were mapped to available functional information from the VitisNet knowledgebase. Of the 2,654 genes, 483 had a functional annotation that was listed as either "Unknown", "Unclear", "Unknown protein", or "No hit." Of those 483 genes, 121 had accompanying information from the Gene Ontology, so a possible associated function could be deduced. For 72 genes no functional or supporting information was available due to those genes being identified by a Vcost.v3 gene name. All clusters were functionally annotated, but the results of the cluster analysis led the focus for further investigation to clusters #2, #3, #13. Figures 24, 25, 26 show the Gene Ontology terms for biological process (requiring at least two differentially expressed genes involved in the process), along with the number of genes associated with each term for clusters of interest.

For cluster #2, 20 of the 42 genes were functionally associated with metabolism. Four other genes were functionally associated with signaling. There were also 3 genes functioning as transcription factors for gene regulation. For the cluster, 31 genes mapped to 24 GO biological process terms. Eight GO terms had at least 2 genes associated to them, with oxidation reduction, metabolic process, and photosynthesis/ light harvesting, having 8 genes, 5 genes, and 4 genes, respectively.

For cluster #3, 26 of the 78 genes were functionally associated with metabolism, including carbohydrate, protein, and lipid. Fourteen genes were associated signaling, including hormone signaling and signaling pathways. Two genes were associated with stress response and 3 genes were from gene families with diverse functions. For the cluster, 48 genes mapped to 75 GO biological process terms, with 16 GO terms mapping to 2 or more genes. Metabolic process, oxidation reduction, and regulation of transcription, DNA-dependent were the top terms, with 7,

6, and 5 genes, respectively. Genes mapped to terms involving subclasses of transcription regulation, which if these terms were combined, transcription regulation would be the term with the most genes.

For cluster #13, 5 of the 18 genes were associated with glutathione metabolism. Three genes were associated with proteolysis, specifically protease inhibition. Two other genes were associated with starch and sucrose metabolism. The function of auxin signaling and the transcript factor WRKY each had 1 gene associated with them. The remaining genes involved stress responses. For the cluster, 15 genes mapped to 12 GO biological terms. Four GO terms had 2 or more genes mapped to them. The terms toxin catabolic process, glutathione metabolic process, and response to wounding had the largest number of genes, with 5, 4, and 3, respectively.

The enrichment analysis reported limited results for all three clusters. For cluster #2, no GO terms were significantly overrepresented for any ontology category. For cluster #3, starch metabolic process was the only biological process term that overrepresented. Two terms, chloroplast stroma and plastid stroma, were significantly overrepresented for the cellular component ontology category. For cluster #13, two biological process terms, response to stress and response to stimulus, were overrepresented. There were also nine molecular function terms involving peptidase activity or protein binding which were overrepresented.

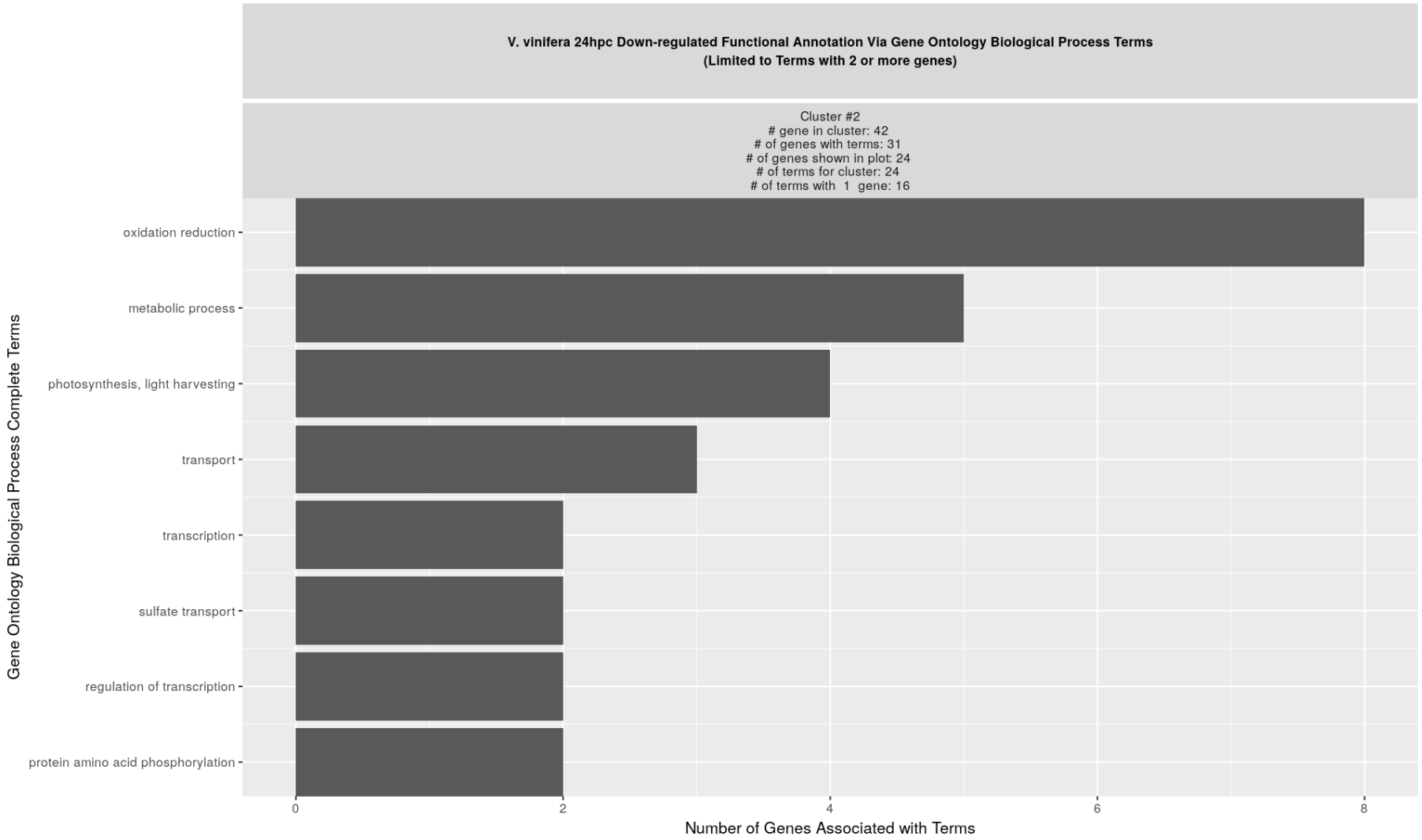


Figure 24: Vitis vinifera - 24hpc down-regulated (cluster #2) biological process GO terms

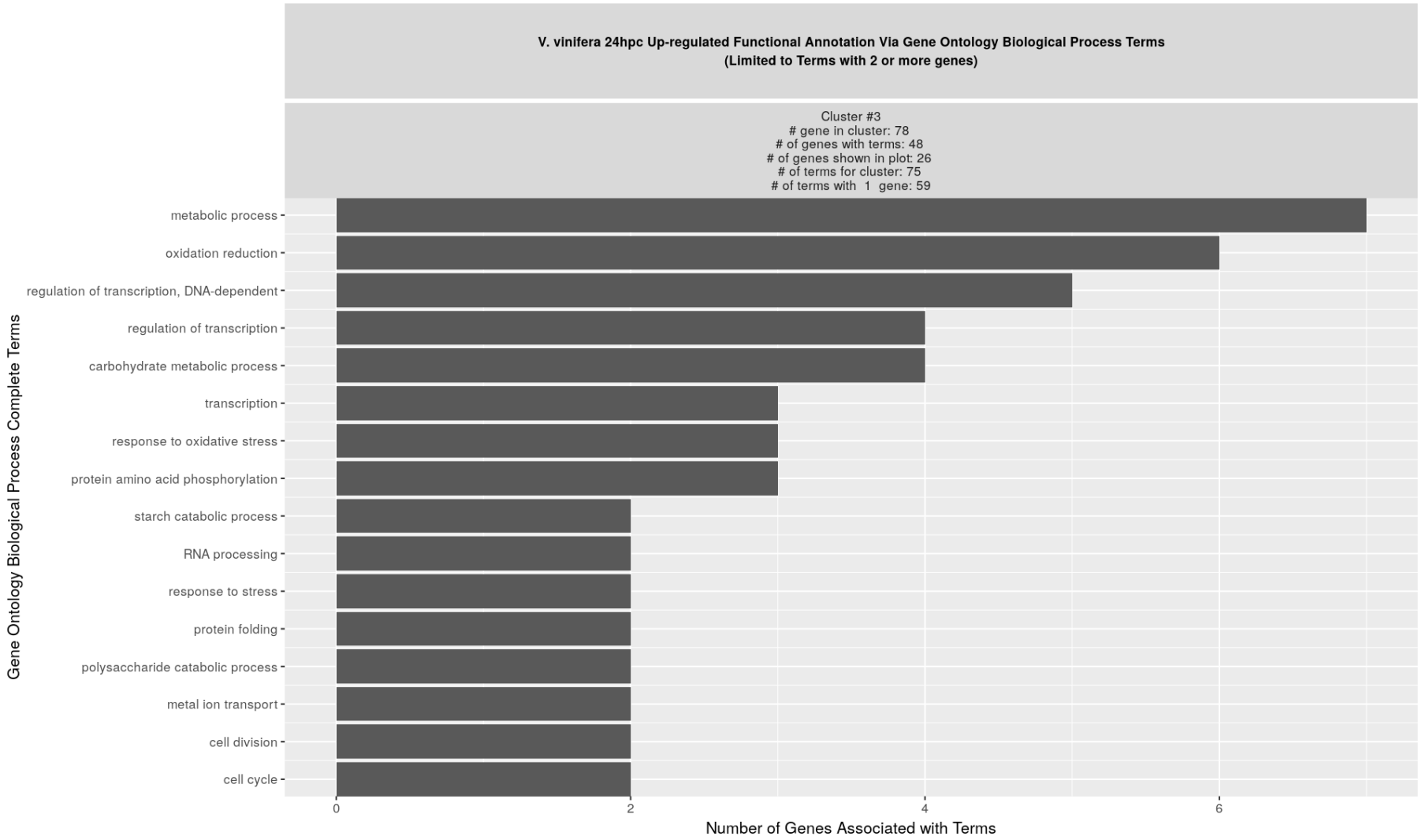


Figure 25: Vitis vinifera - 24hpc up-regulated (cluster #3) biological process GO terms

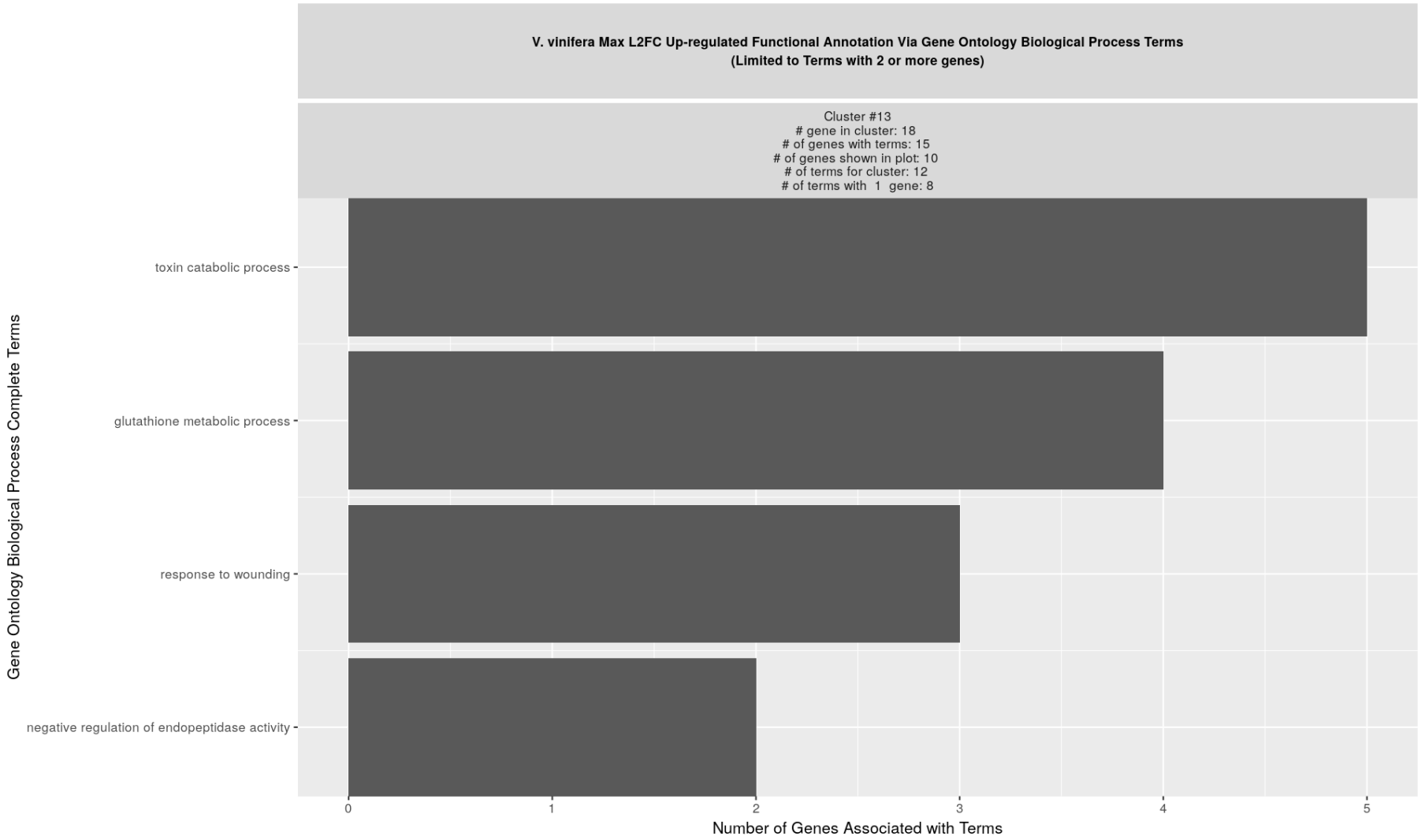


Figure 26: Vitis vinifera – Max up-regulated (cluster #13) biological process GO terms

For the *A. thaliana* dataset, functional information was available from the Gene Ontology for 588 genes. All clusters were functionally annotated, though the focus was on the clusters of interest, based on the expression patterns at 12hpc and 24hpc as described in the cluster analysis section.

Clusters showing up-regulation at 12hpc included cluster #2, #4, and #6. For cluster #2, 54 of the 64 genes mapped to 143 GO terms, and 39 GO terms were associated with 2 or more genes. The top terms included response to wounding with 9 genes, response to abscisic acid with 9 genes, and response to water deprivation with 8 genes. It can be noted that a number of defense response terms are also present with associated gene numbers ranging from 2 to 7. For the 18 genes in cluster #4, 14 mapped to 40 GO terms. Of these terms, only 4 mapped to more than 1 gene. Biological process, response to abscisic acid, and response to cold were the top terms with 5, 3, and 2 genes, respectively. Cluster #6 had 41 of 51 genes that mapped to a total of 85 terms. Seventeen of those terms mapped to at least 2 genes. The top terms for this cluster were regulation of transcription, DNA-templated with 8 genes, biological process with 8 genes, and response to salt stress with 4 genes. Similar to cluster #2, cluster #6 also included some defense response terms.

Cluster #9 showed down-regulation in the 12hpc treatment group. Of the 24 genes in the cluster, 23 genes mapped to a total of 81 GO terms. Of those terms, 16 had 2 or more associated genes. The top terms included response to karrikin and response to auxin with 8 genes each and response to salt stress and regulation of transcription, DNA-templated with 4 genes each.

Clusters showing up-regulation at 24hpc included cluster #4, cluster #9, and cluster #10. The functional information for clusters #4 and #10 are the same as listed above. For cluster #10, 36 of the 45 genes mapped to 90 GO terms. For these terms, 14 terms were associated with

2 or more genes. Two terms, response to cadmium ion and biological process ties for the top number of genes with 4 genes each. Response to cold, protein folding, and chloroplast organization were the next top terms with 3 genes each.

Clusters showing down-regulation at 24hpc included clusters #6, #3, and #11. The functional information for clusters #6 is the same as listed above. For clusters #3, 41 of the 50 genes mapped to 105 GO terms. For these terms, 20 terms were associated with 2 or more genes. Biological process was the top term with 7 genes. Defense response to bacterium was the next largest term with 5 genes. Response to salt stress and protein phosphorylation were the next top terms with 4 genes each. For cluster #11, 51 of the 67 genes mapped to 103 GO terms. Of these terms, 25 terms had 2 more or more genes associated with them. Response to salt stress was the largest term with 7 genes. Regulation of transcription, DNA-templated and biological process were the next largest terms with 5 genes each. Two terms, response to oxidative stress and response to cadmium ion, tied for the next largest with 4 genes apiece.

Figures 27, 28, 29, and 30 show the enriched biological process (BP) terms for the clusters from the AgriGO analyses. The size of the point corresponds to the number of genes associated to each term. The color represents the false discovery rate, following a cutoff off 0.05. The gene ratio, which represents the number of genes associated to the GO relative to the number of total genes in the cluster, shown on the x-axis. There is the general trend that the larger the number of genes associated with a term, the larger the gene ratio.

Figure 31 shows the GO terms for the up and down genes between the PEN1 and Col-0 genotypes. The PEN1 genotype had 82 genes that were up-regulated compared to the Col-0 genotype. Of those genes, 51 genes mapped to 104 GO terms. Thirteen GO terms mapped to 2 or more genes. The top term was biological process with 16 genes. The second largest term was DNA-templated regulation of transcription with 6 genes. The GO terms oxidation-reduction

process and multicellular organism development were the next largest, each with 4 genes associated with them. The PEN1 genotype had 52 genes down-regulated compared to the Col-0 genotype and 37 of those genes had 87 GO terms associated with them. Thirteen of those genes had 2 or more genes mapped to them. The top GO term was biological process with 11 genes. The second largest GO term was defense response with 5 genes. The next largest terms with 3 genes each were the GO terms response to fungus, response to abscisic acid, DNA-templated regulation of transcription, and regulation of stomatal movement.

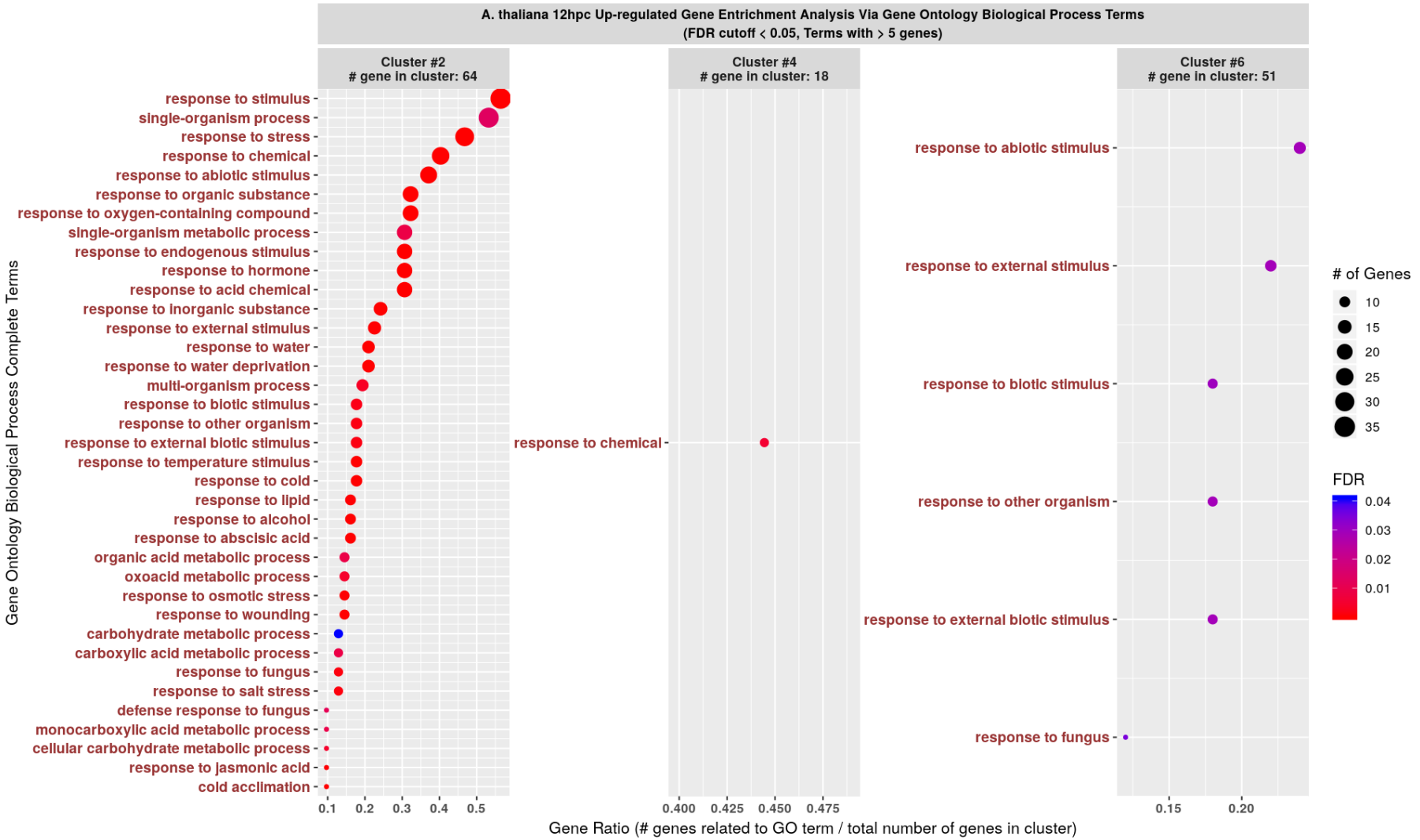


Figure 27: Arabidopsis thaliana - 12hpc up-regulated (clusters #2, #4, #6) – enriched BP GO terms

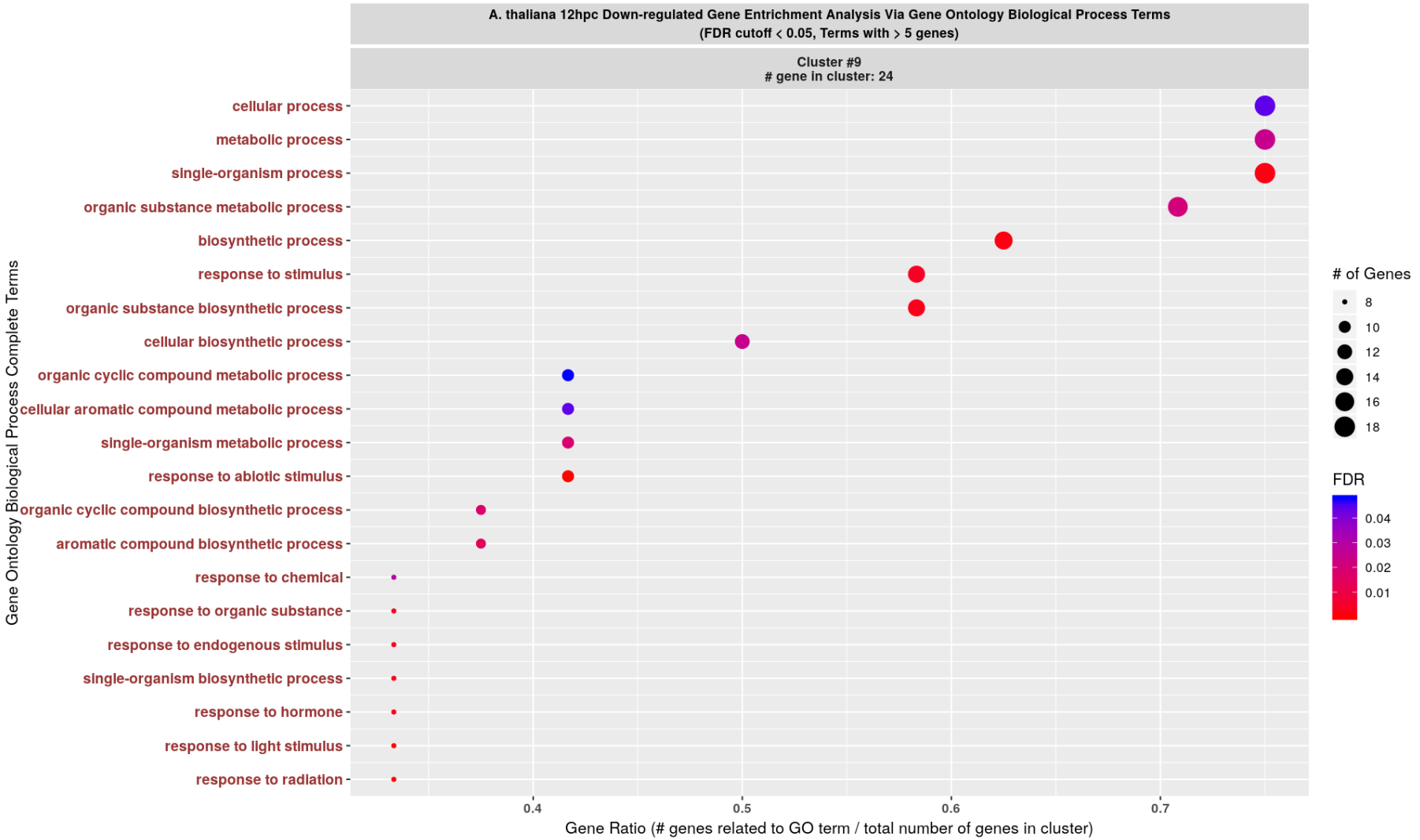


Figure 28: Arabidopsis thaliana - 12hpc down-regulated (cluster #9) – enriched BP GO terms

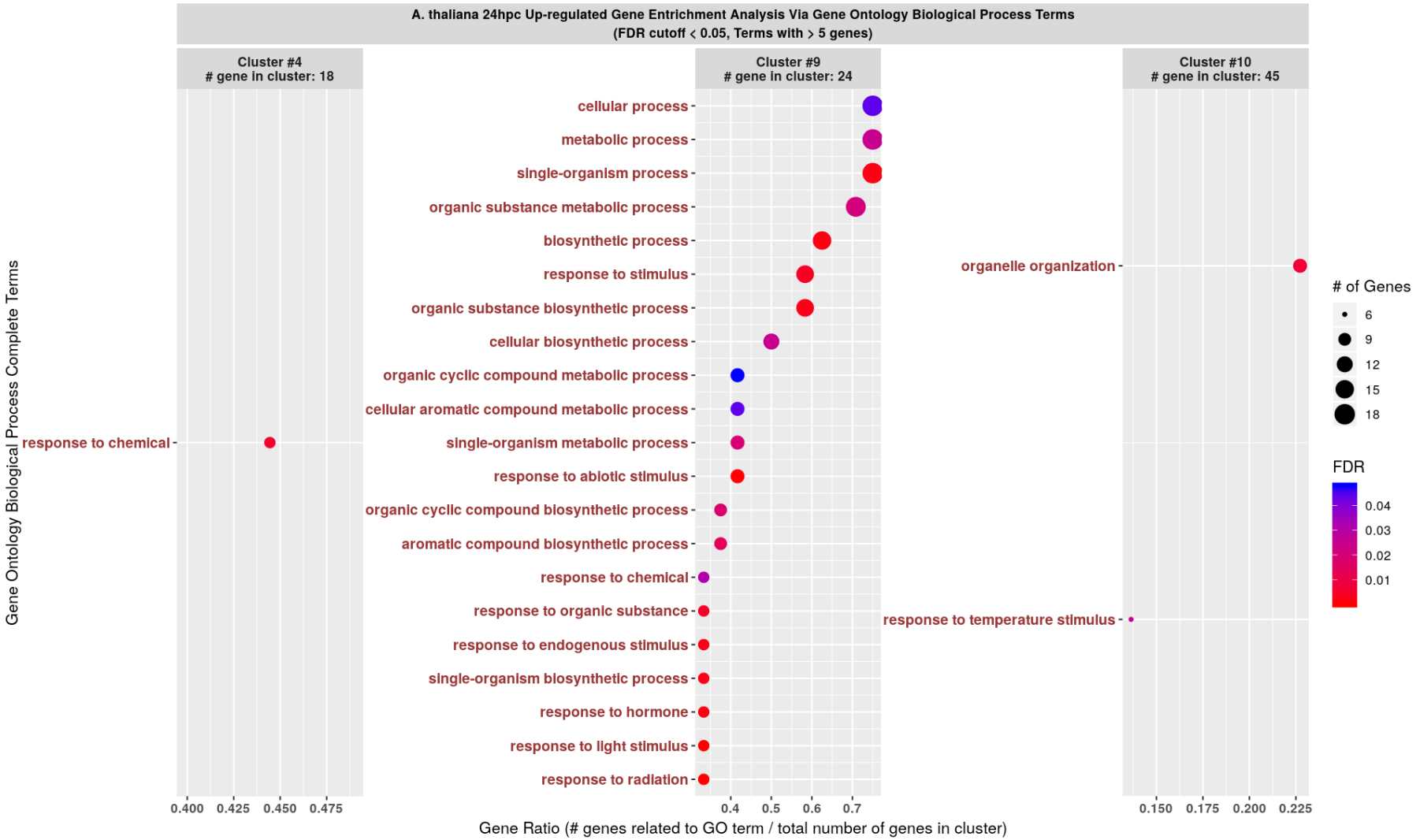


Figure 29: Arabidopsis thaliana - 24hpc up-regulated (clusters #4, #9, #10) – enriched BP GO terms

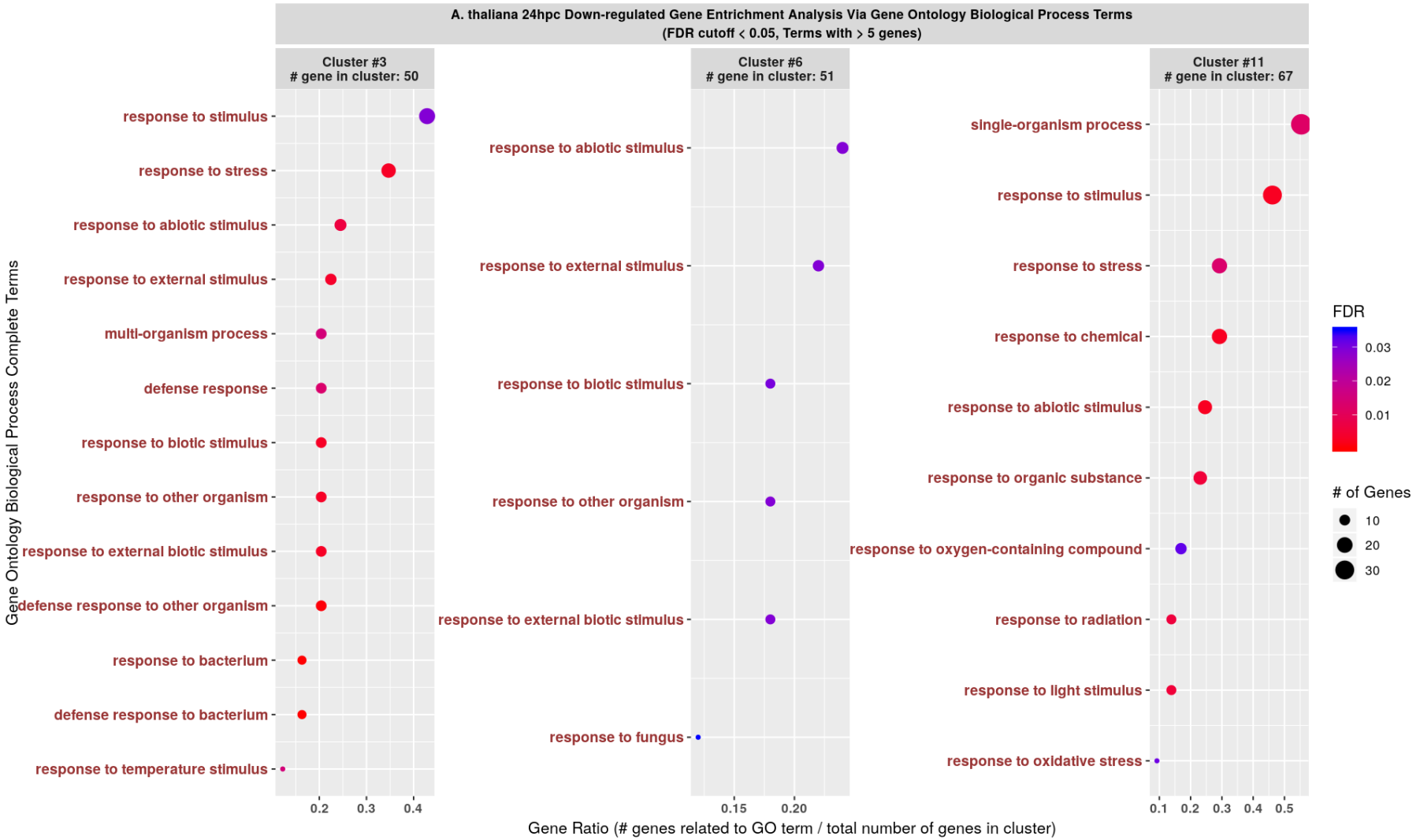


Figure 30: Arabidopsis thaliana - 24hpc down-regulated (clusters #3, #6, #11) – enriched BP GO terms

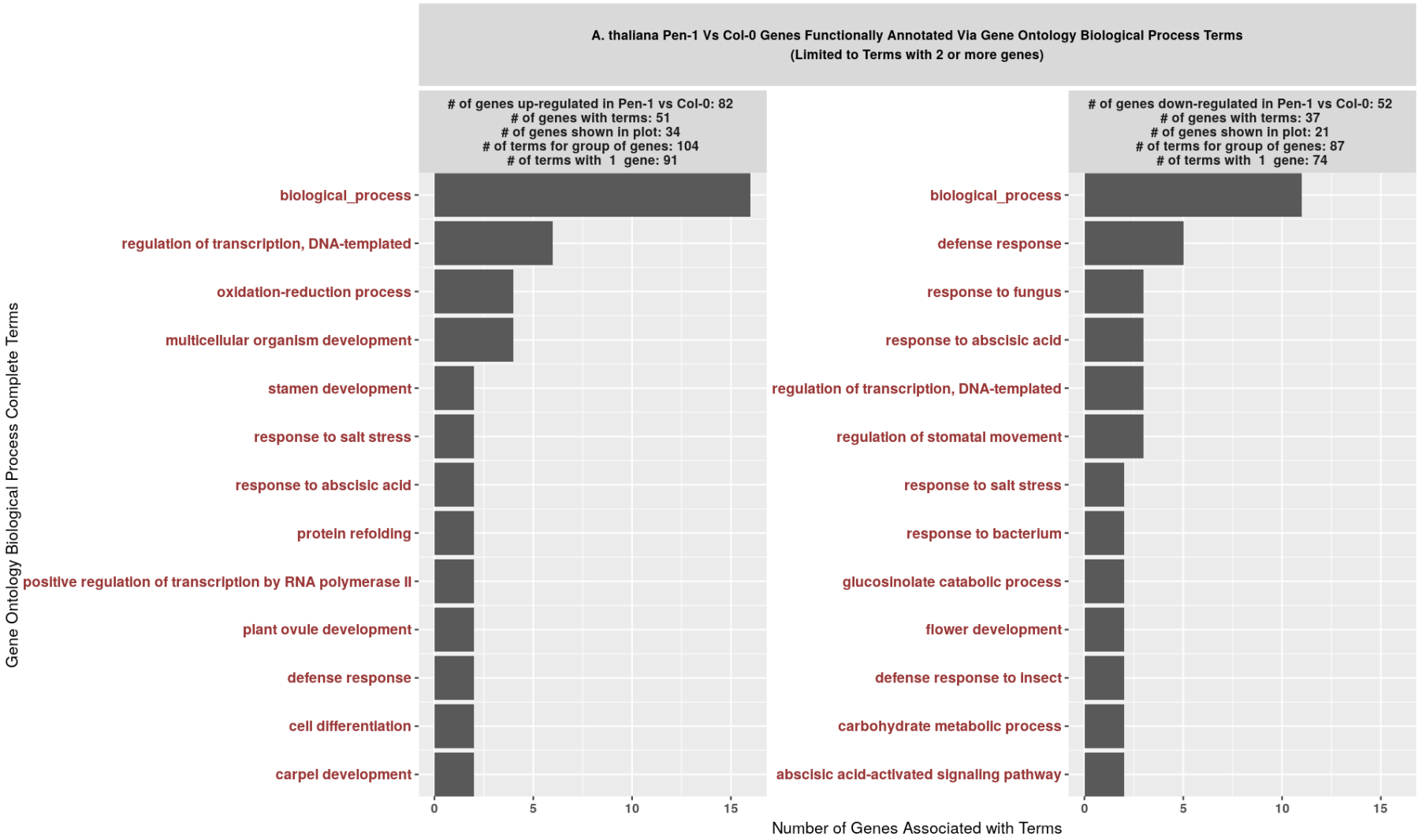


Figure 31: Arabidopsis thaliana Biological Process GO terms for PEN1 compared to Col-0

Pathway Analysis

The clusters for *V. vinifera* were submitted to VitisPathways to determine their associated pathways. Heatmaps were created showing the clusters, ranked by their mean LFC at 24hpc, with the color scales of the heatmap based on the permuted p-value. The significant pathways for the 7 gene clusters with a positive mean LFC are shown in Figure 32. Focusing on cluster #3, circadian rhythm, HSFs, Starch and sucrose metabolism, pseudo ARR-B, Cytokinin signaling, and AP2 EREBP were determined significant. Other clusters with a positive mean LFC shared some, though not all of these pathways.

Figure 32: Up-regulated Pathways for Vitis vinifera Clusters

The significant pathways for the 6 gene clusters with a negative mean LFC are shown in Figure 33. Focusing on cluster #2, photosynthesis related pathways, specifically antenna proteins and transport electron carriers, were down regulated in this cluster. Other clusters shared this down-regulation in photosynthesis activity. The pathway Porters cat 30 to 64 refers to transporter categories 30 through 64. This pathway, along with the transport electron carriers pathway, show that cellular respiration was down-regulated.

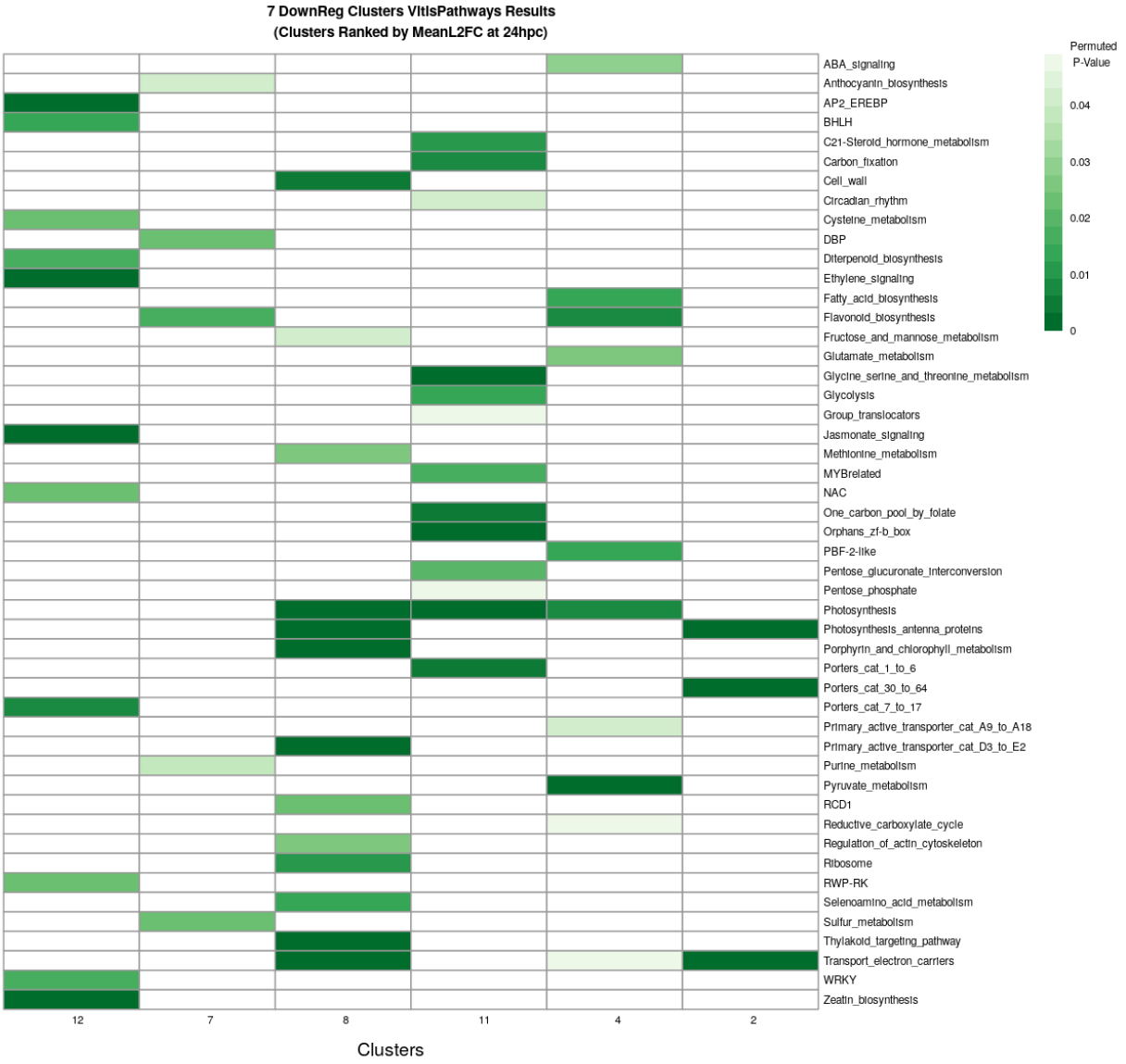


Figure 33: Down-regulated Pathways for Vitis vinifera Clusters

The significant pathways for the 5 gene clusters for *A. thaliana* with an up-regulation at 12hpc are shown in Figure 34. As with the functional annotation, the focus was on the clusters of interest as described above. Custer #2 which had the largest mean LFC at 12hpc, had 5 pathways significantly up-regulated. Two pathways with the lowest permuted p-value were the coumarin biosynthesis pathway and the jasmonic signaling pathway. Coumarin are found in higher plants where they originate from the phenylpropanoid pathway (Bourgaud et al., 2006).

The jasmonic signaling pathway involves a class of plant hormones that play essential roles throughout development, including during plant reproduction, leaf senescence, and in response to many biotic and abiotic stresses (Larrieu & Vernoux, 2016). Cluster #4 showed multiple pathways associated with choline and phosphatidylcholine biosynthesis. As a fundamental metabolite, choline, contributes to the synthesis of the membrane phospholipid, phosphatidylcholine, which accounts for 40 to 60% of lipids in non-plastid plant membranes (Schläpfer et al., 2017). The inosine-5'-phosphate biosynthesis II, purine nucleotides de novo biosynthesis II, and starch degradation II pathways are involved with cellular metabolism. Cluster #6 showed 11 pathways significantly up-regulated. Four pathways involved asparagine or aspartate biosynthesis. According to the PlantCyc database, aspartate and asparagine are involved in protein synthesis and the transportation of nitrogen throughout the plant (Schläpfer et al., 2017). The remaining pathways look to deal with cellular metabolism and secondary metabolites.

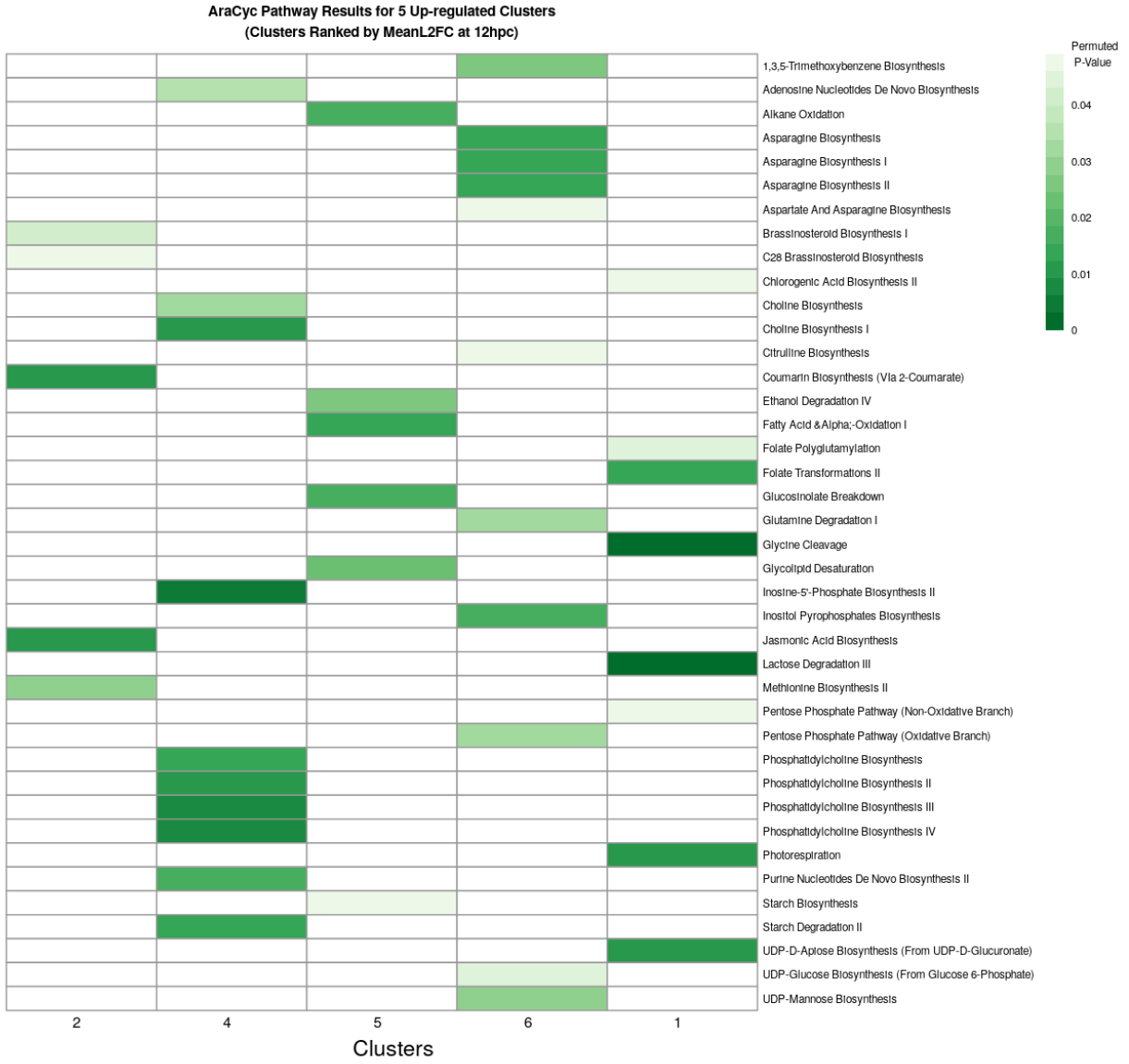


Figure 34: 12hpc up-regulated pathways for Arabidopsis thaliana clusters

The significant pathways for the 8 gene clusters for *A. thaliana* with a down-regulation at 12hpc are shown in Figure 35. Cluster #9 which had the largest negative mean LFC at 12hpc, had 10 pathways significantly down-regulated. Four of these pathways, involve secondary metabolite biosynthesis, such as the flavonoid biosynthesis, phlorizin biosynthesis, and pinobanksin biosynthesis pathways. The resveratrol biosynthesis pathway produces the polyphenol resveratrol which is formed as a general response to biotic and abiotic stresses

(Schläpfer et al., 2017). Four polyamine pathways, including the spermidine biosynthesis I and the spermine biosynthesis pathways, showed up-regulation in cluster #9. As examples of polyamines, spermidine and spermine show involvement in many biological processes, including nucleic acid binding, membrane stabilization, and enzyme stimulation (Schläpfer et al., 2017).

Figure 35: 12hpc down-regulated pathways for Arabidopsis thaliana clusters

The significant pathways for the 7 gene clusters for *A. thaliana* with an up-regulation at 24hpc are shown in Figure 36. The clusters of interest include cluster #4, #9, and #10. Descriptions of the significant pathways for cluster #4 and #9 are the same as above. For cluster #10, ten pathways were shown to be significantly enriched. Two pathways were associated with 4-hydroxybenzoate biosynthesis. 4-hydroxybenzoate is widespread in plants and play a major role in plant defense against pathogens (Caspi et al., 2016). The simple coumarins biosynthesis pathway is associated with the synthesis of secondary metabolites. Three pathways involve either methionine biosynthesis or degradation. The S-adenosyl-L-methionine (SAM) cycle II also showed enrichment. The sucrose degradation VI (anaerobic) pathway contributes to energy and substrate production for plant growth (Schläpfer et al., 2017).

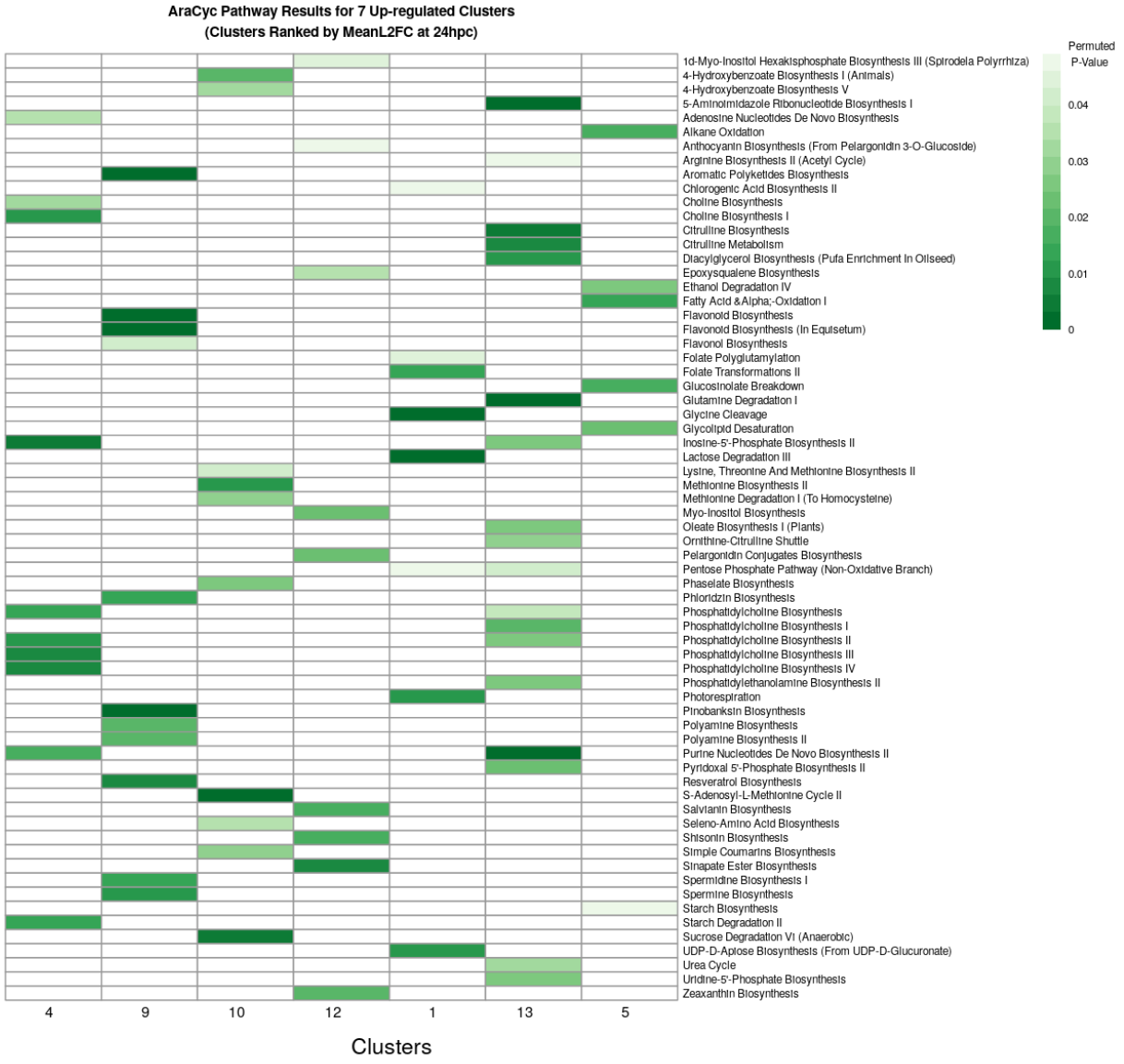


Figure 36: 24hpc up-regulated pathways for Arabidopsis thaliana clusters

The significant pathways for the 6 gene clusters for *A. thaliana* with a down-regulation at 24hpc are shown in Figure 37. The clusters of interest include cluster #6, #3 and #11. Descriptions of the significant pathways for cluster #6 are the same as above. For clusters #3 and #11, ten pathways were shown to be significantly enriched. The four pathways with the lowest permuted p-value showed involvement with either amino acid degradation or amino acid biosynthesis.

Figure 37: 24hpc down-regulated pathways for Arabidopsis thaliana clusters

The significant pathways that showed up and down regulated in the PEN1 genotype compared to the Col-0 genotype are shown in Table 12. Four of the pathways in PEN1 showed associations to the ethylene synthesis. Ethylene synthesis can be induced by environmental stress, as well as other plant growth hormones (Schläpfer et al., 2017). Other pathways showed association to generating energy and precursor metabolites, specifically the pentose phosphate

pathways. Two pathways down-regulated in the PEN1 involve glucosinolate breakdown. Glucosinolates play an important role in plant defense (Schläpfer et al., 2017).

Up-regulated	Down-regulated
Thiosulfate Disproportionation III (rhodanese)	Glucosinolate Breakdown
S-Adenosyl-L-Methionine Biosynthesis	Glucosinolate Breakdown (via Thiocyanate-Forming Protein)
Pentose Phosphate Pathway (oxidative Branch)	Coumarin Biosynthesis (via 2-Coumarate)
Flavin Biosynthesis I (bacteria and Plants)	Tetrapyrrole Biosynthesis I
Methionine Degradation I (to Homocysteine)	Heme Biosynthesis I
Pentose Phosphate Pathway	Superpathway Of Proto- And Siroheme Biosynthesis
S-Adenosyl-L-Methionine Cycle II	Adenosine Nucleotides De Novo Biosynthesis
Ethylene Biosynthesis I (plants)	

Table 12: Significant pathways for PEN1 genotype

Organism Comparison

For *V. vinifera*, the significant genes were mapped to the best matching *A. thaliana* gene homolog. Table 13 shows the number of homologs mapped to the significant genes for each cluster.

V. vinifera

Cluster	Significant Genes	Homologs
2	42	34
3	78	64
13	18	9

A. thaliana

Cluster	Significant Genes	Homologs
2	64	66
3	50	88
4	18	24
6	51	61
9	24	76
10	45	59
11	67	58

Table 13: Gene homolog counts for clusters, by organism

Some of the genes from the *V. vinifera* clusters did not have any available homologs as so were underrepresented in the preceding pathways analysis. Some of the genes from the *A. thaliana*

clusters mapped to multiple *V. vinifera* genes. Table 14 shows the intersection between the *V. vinifera* genes with *A. thaliana* homologs and the *A. thaliana* dataset, limited to the clusters of interest for each dataset. Two *A. thaliana* homologs, AT3G52180 and AT5G24470, were found to be consistent between the *V. vinifera* cluster #3 and *A. thaliana* clusters #2, #4, and #6. One *A. thaliana* homolog, AT4G17090, was found to intersect between the *V. vinifera* cluster #3 and *A. thaliana* clusters #4, #9, and #10.

		Significant Genes from *Arabidopsis thaliana* Dataset	
		12hpc	24hpc
		Up-Regulated (Cluster #2, #4, #6)	Up-Regulated (Cluster #4, #9, #10)
Vitis vinifera Genes with *Arabidopsis thaliana* homologs	Up-regulated (Cluster #3)	2	1
		Down-regulated (Cluster #9)	Down-regulated (Cluster #3, #6, #11)
	Down-regulated (Cluster #2)	0	0

Table 14: Vitis vinifera Homologs Intersected with Arabidopsis thaliana Dataset by Cluster of Interest

This intersection was consistent between the *A. thaliana* genes with *V. vinifera* homologs and the *V. vinifera* dataset. A pathways analysis was performed for the homologs for each organism.

The AraCyc pathway was used for the *A. thaliana* homologs of the significant *V. vinifera* genes. Cluster #2 from the *V. vinifera* dataset showed several pathways involved with the biosynthesis of flavonoids. Two pathways, ricinoleate biosynthesis pathway and the vestitol and sativan biosynthesis pathway, were also shown as enriched. These two pathways have associations with plant defense response (Caspi et al., 2016). Cluster #3 from the *V. vinifera* dataset showed multiple pathways associated with the biosynthesis of choline and its phospholipid product, phosphatidylcholine. The spermidine hydroxycinnamic acid conjugates biosynthesis pathway and the glycine biosynthesis II pathway also showed significant enrichment.

The VitisNet pathway was used for the *V. vinifera* homologs from the significant *A. thaliana* genes. Focusing on the *A. thaliana* clusters up-regulated at 12hpc, there were several pathway results that overlapped with the *V. vinifera* pathways up-regulated at 24hpc. These overlapping pathways include cytokinin signaling, circadian rhythm, and pseudoARR-B. The auxin and WRKY pathways also showed enrichment in *A. thaliana*, though these overlaps were not with the 24hpc up-regulated cluster in *V. vinifera* (cluster #3), they did overlap with other up-regulated *V. vinifera* clusters. Cluster #9, the down-regulated 12hpc *A. thaliana* cluster, showed enrichment for flavonoid biosynthesis, phenylpropanoid biosynthesis, and transporter categories 1 to 6. These results did not overlapped any of the *V. vinifera* pathways, however, they were consistent with the AraCyc pathway results for this cluster. For the *A. thaliana* clusters up-regulated at 24hpc, there looked to be some overlap with the methionine metabolism and methane metabolism pathways. The *A. thaliana* clusters down-regulated at 24hpc overlapped with the 24hpc down-regulated cluster in *V. vinifera* (cluster #2) showing enrichment of the photosynthesis pathway, indicating a down-regulation of the pathway.

Discussion

Three hypotheses by Moyer et al. (2016) were introduced at the beginning of this project to explain the cold-SIDR response. The first hypothesis involved the cold temperature impacting photosynthesis efficiency. The second hypothesis is that basic physiological responses to cold may impart some form of disease resistance. The third hypothesis involved the regulation of plant hormones signaling impacting plant cell growth and elongation. The results from this project show that the two datasets show mixed support for these hypotheses. This project also produced additional genes and pathways which could confer the cold-SIDR resistance.

Hypothesis #1

Regarding the first hypothesis, the *V. vinifera* dataset showed support. Genes in cluster #2 showed patterns of down-regulation at 24hpc. Based on the functional information for the genes, metabolism and photosynthesis looked to be down-regulated. This is also supported by the results from the pathways analysis which identified the photosynthesis related pathways, specifically antenna proteins and transport electron carriers being down-regulated for this cluster. Other clusters showing down-regulation at 24hpc (Figure 33) were also associated with photosynthetic activity. This is consistent with the expectation that cold temperature would lead to decreased levels of photosynthetic activity.

The *A. thaliana* dataset showed support for this hypothesis as well. Functional and pathway results for the clusters of interest show an increases in starch and carbohydrate breakdown. While not directly implicated in the photosynthesis pathway, starch and carbohydrate breakdown fuels plant metabolism and growth when they are unable to photosynthesize (Streb & Zeeman, 2012).

Hypothesis #2

Regarding hypothesis number two, the *V. vinifera* dataset showed some support. In cluster #3, no genes were directly labeled as cold response (COR) genes, but 18 genes had a *A. thaliana* homolog characterized as a COR gene. Two specific homologs, COR27 and COR28, showed a range of up-regulation in all of the treatment group, but showed their highest up-regulation in the 24hpc treatment group. As photosynthesis is down-regulated, reactive oxygen species (ROS) accumulate in the cells. These ROS can act as signaling molecules in the response pathways. Evidence for the presence of these ROS is present in cluster #3. Genes mapped to GO terms involving oxidation reduction and response to oxidative stress. Closer inspection showed up-regulation of peroxisome activity.

Results from the *A. thaliana* dataset showed several cold response (COR) genes with up-regulation during the 12hpc and 24hpc treatments. One COR gene of particular interest would be COR413PM1 (AT2G15970) which codes a cold regulate plasma membrane protein. Adherence of the plasma membrane to the cell wall was shown to impact invasion success in other biotrophic fungi (Mellersh & Heath, 2001). The exact function of COR413PM1 in conferring stress tolerance is currently unknown. The protein has homologs which are implicated in conferring osmotic stress tolerance, in addition to cold tolerance (Garwe, Thomson, & Mundree, 2003). Evidence for the presence of these ROS is also present in the *A. thaliana* dataset, most noteworthy in cluster #2. Twenty of the 64 genes showed enrichment for response to an oxygen-containing compound. Clusters #3 and #11 were also enriched for gene associated with response to oxidative stress at 24hpc.

Hypothesis #3

The third hypothesis involved the regulation of plant hormone signaling, impacting plant growth. The *V. vinifera* dataset showed mixed support for this hypothesis as well. Genes associated with the Gibberellin biosynthesis pathway showed down-regulation. The review by Moyer et al. (2016) mentioned that abscisic acid (ABA) could be responsible for the cold-SIDR response. Of the *V. vinifera* genes associated with ABA, 10 showed up-regulation and 13 genes were down-regulated. The expression strength of these genes was rather low, with a LFC ranging from 0.5 to -0.6 and most having between 0.2 and -0.1. Some genes showing up-regulation had no characterized function, however, their *Arabidopsis* orthologs were involved with leaf senescence and response to ABA.

The *A. thaliana* results provided more solid evidence for this hypothesis, specifically the 12hpc enrichment results for cluster #2 (Figure 26). Fifteen of the 64 genes associated with a response to ABA were significantly overrepresented.

Additional Considerations

This project also produced additional findings that were not included in the three hypotheses. First, the jasmonic acid (JA) pathway showed up-regulation in both *V. vinifera* and *A. thaliana* datasets, occurring in the clusters of interest to cold-SIDR. Classically, salicylic acid (SA) signaling is associated with biotroph resistance, while JA signaling plays an important role in necrotrophic pathogen resistance (Duan et al., 2014). In most cases of dicotyledonous plants, the JA and SA signaling pathways are expected to interact antagonistically (Tamaoki et al., 2013). This generalization is disputed in grapevine as JA signaling has been implicated in host resistance against biotrophs (Belhadj et al., 2006) (Guerreiro et al., 2016).

Plants have evolved with a variety of responses to cope with both abiotic and biotic stress. The interaction between biotic and abiotic stress responses is orchestrated by hormone signaling pathways that may induce or antagonize one another (Atkinson & Urwin, 2012). Cold-SIDR is an example of one of such response. Transcription factors, ROS, and other signaling responses are key components of this pathway crosstalk. Several of these key components were expressed and up-regulated in the clusters of interest in both organisms. The challenging element is that even at low expression levels, these signaling components can initiate cascades leading to many genes being highly expressed.

The role of PEN1 in the cold-SIDR response remains unclear at a molecular level. Though no genes with a documented association with PEN1 were differentially expressed, additional genes of interest are reflected in genes also down-regulated in the PEN1 genotype compared to the Col-0. The second largest negative $\log_2$ fold-change was the gene CNGC12 (AT2G46450). CNGC12 has been related to defense response and acts as a positive regulator of resistance against avirulent fungal pathogen (Swarbreck et al., 2008). CNGC12 has been found to play a significant role in Ca_2^+ signaling, which leads to the mediation of several physiological

processes, including senescence (Abdel-Hamid, Chin, Moeder, & Yoshioka, 2011). The third largest negative log_2 fold-change was the gene, QQS (AT3G30720), which appears to modulate carbon/nitrogen allocation in *Arabidopsis* (Swarbreck et al., 2008). Recently, QQS has been linked with increased resistance to viruses, bacteria and fungi in Arabidopsis and soybean (Qi et al., 2019). Further investigation is needed to understand the connections between PEN1 and these additional down-regulated genes.

Future Directions

Further research should be done to investigate if the gene expression profiles vary depending on the time of day that the cold event occurs. This could be achieved by performing a project similar to this one but invert the time of day that the cold treatments were applied to the 12hpc, 24hpc, 36hpc, and 48hpc treatments. This would assist in determining if the cold-SIDR response is a function of the cold response genes or if it is rooted in genes responsible for circadian rhythm. Another focus would be investigating cold-SIDR with methods that account for alternative splicing. Although not highly characterized in plants, it is likely that alternative splicing can lead to major changes in response to both abiotic and biotic stresses (Calixto et al., 2018). Lastly, the continued development of pathway architectures involved with stress response is necessary to elucidate the mechanism of cold-SIDR.

Conclusion

Cold-SIDR is a phenomenon resulting in decrease susceptibility to biotroph infection following exposure to an acute (less than 4 hours) cold (below 8°C) event. This phenomenon has been observed in the species *V. vinifera* and *A. thaliana*, as well as other plant species. The exact genetic mechanism of cold-SIDR remains unknown. The results of this project provide support for the several hypotheses for explaining the cold-SIDR response. Understanding the mechanism behind cold-SIDR would allow for improved disease prediction models. With more accurate models, the total amount of fungicides applied could be reduced while simultaneously preventing severe disease development.

References

Abdel-Hamid, H., Chin, K., Moeder, W., & Yoshioka, K. (2011). High throughput chemical screening supports the involvement of Ca2+ in cyclic nucleotide-gated ion channel-mediated programmed cell death in Arabidopsis. Plant Signaling & Behavior, 6(11), 1817–1819. https://doi.org/10.4161/psb.6.11.17502.

Anders, S., Pyl, P. T., & Huber, W. (2015). HTSeq—a Python framework to work with high-throughput sequencing data. Bioinformatics, 31(2), 166–169. https://doi.org/10.1093/bioinformatics/btu638.

Andrews S. (2010). FastQC: a quality control tool for high throughput sequence data. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc.

Atkinson, N. J., & Urwin, P. E. (2012). The interaction of plant biotic and abiotic stresses: from genes to the field. Journal of Experimental Botany, 63(10), 3523–3543. https://doi.org/10.1093/jxb/ers100.

Bendek, C. E., Campbell, P. A., Torres, R., Donoso, A., & Latorre, B. A. (2007). The risk assessment index in grape powdery mildew control decisions and the effect of temperature and humidity on conidial germination of Erysiphe necator. Spanish Journal of Agricultural Research, 5(4), 522–532.

Berardini, T. Z., Reiser, L., Li, D., Mezheritsky, Y., Muller, R., Strait, E., & Huala, E. (2015). The arabidopsis information resource: Making and mining the "gold standard" annotated reference plant genome. Genesis, 53(8), 474–485. https://doi.org/10.1002/dvg.22877.

Bourgaud, F., Hehn, A., Larbat, R., Doerper, S., Gontier, E., Kellner, S., & Matern, U. (2006). Biosynthesis of coumarins in plants: a major pathway still to be unravelled for cytochrome P450 enzymes. Phytochemistry Reviews, 5(2), 293–308. https://doi.org/10.1007/s11101-006-9040-2.

Calixto, C. P. G., Guo, W., James, A. B., Tzioutziou, N. A., Entizne, J. C., Panter, P. E., … Brown, J. W. S. (2018). Rapid and Dynamic Alternative Splicing Impacts the Arabidopsis Cold Response Transcriptome. The Plant Cell, 30(7), 1424–1444. https://doi.org/10.1105/tpc.18.00177.

Canaguier, A., Grimplet, J., Di Gaspero, G., Scalabrin, S., Duchêne, E., Choisne, N., … Adam-Blondon, A.-F. (2017). A new version of the grapevine reference genome assembly (12X.v2) and of its annotation (VCost.v3). Genomics Data, 14, 56–62. https://doi.org/10.1016/j.gdata.2017.09.002.

Caspi, R., Billington, R., Ferrer, L., Foerster, H., Fulcher, C. A., Keseler, I. M., … Karp, P. D. (2016). The MetaCyc database of metabolic pathways and enzymes and the

BioCyc collection of pathway/genome databases. Nucleic Acids Research, 44(D1), D471–D480. https://doi.org/10.1093/nar/gkv1164.

Chen, S. (2018, December). Question: what cause poly-G from NextSeq [Blog post]. Retrieved from https://www.biostars.org/p/294612/.

Cheng, C.-Y., Krishnakumar, V., Chan, A. P., Thibaud-Nissen, F., Schobel, S., & Town, C. D. (2017). Araport11: a complete reannotation of the Arabidopsis thaliana reference genome. The Plant Journal, 89(4), 789–804. https://doi.org/10.1111/tpj.13415.

Conesa, A., Madrigal, P., Tarazona, S., Gomez-Cabrero, D., Cervera, A., McPherson, A., … Zhang, X. (2016). A survey of best practices for RNA-seq data analysis. Genome Biology, 17(1), 13.

Duan, Z., Lv, G., Shen, C., Li, Q., Qin, Z., & Niu, J. (2014). The role of jasmonic acid signalling in wheat (Triticum aestivum L.) powdery mildew resistance reaction. European Journal of Plant Pathology, 140(1), 169–183. https://doi.org/10.1007/s10658-014-0453-2.

Futschik M, Carlisle B (2005). "Noise robust clustering of gene expression time-course data." *Journal of Bioinformatics and Computational Biology*, 965-988. http://mfuzz.sysbiolab.eu.

Fuller, K. B., Alston, J. M., & Sambucci, O. S. (2014). The value of powdery mildew resistance in grapes: evidence from California. Wine Economics and Policy, 3(2), 90–107.

Gadoury, D. M., Cadle-Davidson, L., Wilcox, W. F., Dry, I. B., Seem, R. C., & Milgroom, M. G. (2012). Grapevine powdery mildew (Erysiphe necator): a fascinating system for the study of the biology, ecology and epidemiology of an obligate biotroph. *Molecular Plant Pathology*, *13*(1), 1–16.

Garwe, D., Thomson, J. A., & Mundree, S. G. (2003). Molecular characterization of XVSAP1, a stress-responsive gene from the resurrection plant Xerophyta viscosa Baker1. Journal of Experimental Botany, 54(381), 191–201. https://doi.org/10.1093/jxb/erg013.

The Gene Ontology Consortium. (2019). The Gene Ontology Resource: 20 years and still GOing strong. Nucleic Acids Research, 47(D1), D330–D338. https://doi.org/10.1093/nar/gky1055.

George, N. I., Bowyer, J. F., Crabtree, N. M., & Chang, C.-W. (2015). An Iterative Leave-One-Out Approach to Outlier Detection in RNA-Seq Data. PLoS ONE, 10(6). https://doi.org/10.1371/journal.pone.0125224.

Guerreiro, A., Figueiredo, J., Sousa Silva, M., & Figueiredo, A. (2016). Linking Jasmonic Acid to Grapevine Resistance against the Biotrophic Oomycete Plasmopara viticola. Frontiers in Plant Science, 7. https://doi.org/10.3389/fpls.2016.00565.

Gururani, M. A., Venkatesh, J., Upadhyaya, C. P., Nookaraju, A., Pandey, S. K., & Park, S. W. (2012). Plant disease resistance genes: Current status and future directions. Physiological and Molecular Plant Pathology, 78, 51–65. https://doi.org/10.1016/j.pmpp.2012.01.002.

Kegley, S.E., Hill, B.R., Orme S., Choi A.H., PAN Pesticide Database, Pesticide Action Network, North America (Oakland, CA, 2016), http://www.pesticideinfo.org/DS.jsp?sk=29141.

Kolde, R. (2019). pheatmap: Pretty Heatmaps (Version 1.0.12). Retrieved from https://CRAN.R-project.org/package=pheatmap.

Korpelainen, E., Tuimala, J., Somervuo, P., Huss, M., & Wong, G. (2014). RNA-seq data analysis: a practical approach. CRC Press.

Larrieu, A., & Vernoux, T. (2016). Q&A: How does jasmonate signaling enable plants to adapt and survive? BMC Biology, 14(1), 79. https://doi.org/10.1186/s12915-016-0308-8.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., … 1000 Genome Project Data Processing Subgroup. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics (Oxford, England), 25(16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352.

Li, J., & Tibshirani, R. (2013). Finding consistent patterns: A nonparametric approach for identifying differential expression in RNA-Seq data. Statistical Methods in Medical Research, 22(5), 519–536. https://doi.org/10.1177/0962280211428386.

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology, 15, 550. https://doi.org/10.1186/s13059-014-0550-8.

Mellersh, D. G., & Heath, M. C. (2001). Plasma Membrane –Cell Wall Adhesion Is Required for Expression of Plant Defense Responses during Fungal Penetration. The Plant Cell, 13(2), 413–424.

Micali, C., Göllner, K., Humphry, M., Consonni, C., & Panstruga, R. (2008). The Powdery Mildew Disease of Arabidopsis: A Paradigm for the Interaction between Plants and Biotrophic Fungi. The Arabidopsis Book / American Society of Plant Biologists, 6, e0115. http://doi.org/10.1199/tab.0115.

Moyer, M. (2011). Understanding Grapevine Powdery Mildew In New York State: Biology, Epidemiology And Risk Assessment. Retrieved from http://ecommons.cornell.edu/handle/1813/33581.

Miura, K., & Furumoto, T. (2013). Cold Signaling and Cold Response in Plants. International Journal of Molecular Sciences, 14(3), 5312–5337. https://doi.org/10.3390/ijms14035312.

Moyer, M. M., Gadoury, D. M., Cadle-Davidson, L., Dry, I. B., Magarey, P. A., Wilcox, W. F., & Seem, R. C. (2010). Effects of acute low-temperature events on development of Erysiphe necator and susceptibility of Vitis vinifera. Phytopathology, 100(11), 1240–1249.

Martinson, T., & Wilcox, W. F. (2013). Powdery Mildew: How Important is Overwintering Inoculum? Appellation Cornell, (15). Retrieved from https://grapesandwine.cals.cornell.edu/newsletters/appellation-cornell/2013-newsletters/issue-15/grapes-101.

Moyer, M. M., Londo, J., Gadoury, D. M., & Cadle-Davidson, L. (2016). Cold Stress-Induced Disease Resistance (SIDR): indirect effects of low temperatures on host-pathogen interactions and disease progress in the grapevine powdery mildew pathosystem. European Journal of Plant Pathology, 144(4), 695–705.

Osier, M. V. (2016). VitisPathways: gene pathway analysis for *V. vinifera*. VITIS - Journal of Grapevine Research, 55(3), 129–133. https://doi.org/10.5073/vitis.2016.55.129-133.

Pessina, S., Lenzi, L., Perazzolli, M., Campa, M., Dalla Costa, L., Urso, S., … Malnoy, M. (2016). Knockdown of MLO genes reduces susceptibility to powdery mildew in grapevine. Horticulture Research, 3, 16016. https://doi.org/10.1038/hortres.2016.16.

Petrov, V., Hille, J., Mueller-Roeber, B., & Gechev, T. S. (2015). ROS-mediated abiotic stress-induced programmed cell death in plants. Frontiers in Plant Science, 6. https://doi.org/10.3389/fpls.2015.00069.

Qi, M., Zheng, W., Zhao, X., Hohenstein, J. D., Kandel, Y., O'Conner, S., … Li, L. (2019). QQS orphan gene and its interactor NF-YC4 reduce susceptibility to pathogens and pests. Plant Biotechnology Journal, 17(1), 252–263. https://doi.org/10.1111/pbi.12961.

Qiu, W., Feechan, A., & Dry, I. (2015). Current understanding of grapevine defense mechanisms against the biotrophic fungus (Erysiphe necator), the causal agent of powdery mildew disease. Horticulture Research, 2, 15020. https://doi.org/10.1038/hortres.2015.20.

Quinlan AR and Hall IM, 2010. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 26, 6, pp. 841–842.

R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, https://www.R-project.org/.

Raanan, R., Gunier, R. B., Balmes, J. R., & Beltran, A. J. (2017). Elemental sulfur use and associations with pediatric lung function and respiratory symptoms in an agricultural community. Environmental Health Perspectives, 125(8). https://doi.org/10.1289/EHP528.

Schläpfer, P., Zhang, P., Wang, C., Kim, T., Banf, M., Chae, L., … Rhee, S. Y. (2017). Genome-Wide Prediction of Metabolic Enzymes, Pathways, and Gene Clusters in Plants. Plant Physiology, 173(4), 2041–2059. https://doi.org/10.1104/pp.16.01942.

Singh, R., Ming, R., & Yu, Q. (2016). Comparative Analysis of GC Content Variations in Plant Genomes. Tropical Plant Biology, 9, 136–149. https://doi.org/10.1007/s12042-016-9165-4.

Streb, S., & Zeeman, S. C. (2012). Starch Metabolism in Arabidopsis. The Arabidopsis Book / American Society of Plant Biologists, 10. https://doi.org/10.1199/tab.0160.

Tamaoki, D., Seo, S., Yamada, S., Kano, A., Miyamoto, A., Shishido, H., … Gomi, K. (2013). Jasmonic acid and salicylic acid activate a common defense system in rice. Plant Signaling & Behavior, 8(6). https://doi.org/10.4161/psb.24260.

Tandonnet, S., & Torres, T. T. (2016). Traditional versus 3′ RNA-seq in a non-model species. Genomics Data, 11, 9–16. https://doi.org/10.1016/j.gdata.2016.11.002.

Tian, T., Liu, Y., Yan, H., You, Q., Yi, X., Du, Z., … Su, Z. (2017). agriGO v2.0: a GO analysis toolkit for the agricultural community, 2017 update. Nucleic Acids Research, 45(Web Server issue), W122–W129. https://doi.org/10.1093/nar/gkx382.

UC-IMP (University of California- Integrated Pest Management Program). (2008, November). Powdery Mildew on Vegetables. Retrieved from http://ipm.ucanr.edu/PMG/PESTNOTES/pn7406.html.

U.S. Department of Commerce. (2017, March 6). 2015 Annual Retail Trade Report. Retrieved from https://www.census.gov/retail/index.html.

Wang, Z., Gerstein, M., & Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. Nature Reviews. Genetics, 10(1), 57–63. https://doi.org/10.1038/nrg2484.

Wickham, H. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2016.

Wilcox, W. F. (2003). Grapevine powdery mildew. Retrieved 9/15/17 From https://ecommons.cornell.edu/handle/1813/43121.

WineAmerica. (2014). About the United States Wine and Grape Industry. Retrieved from http://wineamerica.org/policy/by-the-numbers.

WineAmerica. (2017). Economic Impact Report. Retrieved from http://wineamerica.org/impact.

Youakim, S. (2006). Occupational health risks of wine industry workers. British Columbia Medical Journal, 48(8), 386.